Modeling Melodic Dictation

David John Baker

2018-10-12

Contents

Chapter 1

Significance of the Study

All students pursing a Bachelor's degree in Music from universities accredited by the National Association of Schools of Music must learn to take melodic dictation (Nat, 2018, Section VIII.6.B.2.A). Melodic dictation is a cognitively demanding process that requires students to listen to a melody, retain it in memory, and then use their knowledge of Western musical notation in order to recreate the mental image of the melody on paper in a limited time frame. As of 2018 there are 647 Schools of Music belonging to National Association of Schools of Music (NASM) CITE WEBSITE, meaning that hundereds of students every year will be expected to learn this challenging task as part of their Aural Skills education. The logic being that as one improves in their abiliy to take melodic dictation, this practice of critical and active listening develops as a means to improve one's ability to "think in music" and thus become a more compotent musician. While learning Aural Skills has been a hallmark of being educated within the Western conservatory tradition, the rationale behind both the how and why of aural skills is often thought of as being esoteric. Throughout the past century, people have disagreed on exactly how one does go about learning a melody with different areas of research each attacking the problem from a different angle.

Despite its ubiqiquity in curricula within School of Music settings, research on topics pertain to how aural skills are acquired is limited at best. [Citations here about the cosntant calls butler, klondoski, pembrook] The fields of music theory and cognitive psychology are best positioned to make progress on this question, but often the skills required to be well versed ein ither of these subjects are disparate, published in other journals, and the research with overlap is scarce. This problem is not new and there have been repeated attempts to bridge the gap between practioners of aural skills and people in cognitive psycholgy CITES. Literature from music theory has establisehd conceptual frameworks regarding aural skills Karpinski (2000) and the relavint cognitive psychology literature has explored factors that might contribute to melodic perception (SCHMUKLER SYNERR 2016 2016), and there exists applied literature from the world of music education (CITES).

However, despite these siloed areas of research, we as music researchers do not have an a concrete understanding of exaclty what contributes to HOW individuals learn melodies (HALPERNBARLETT2010). This is peculiar since "how does one learn a melody" seems to be one of the fundamental questions to the fields of music theory, music psychology, as well as music education. Given this lack of understanding, it becomes even more peculiar that this lack of convergence of evidence is then unable to provide a solid baseline as to what student in their aural skills classrooms can be expected to do. (Also something about we should really know this if we are going to grade people on this ability). While no single dissertation can solve any problem completely, this dissertation aims to fill the gap in the literature between aural skills practitioners (theorists and educators) and music psychologists in order to reach conclusion that can be applied systematically in pedagogical contexts. In order to do this I draw both literatures (music and science) in order to demonstrate how tools from both cognitive psychology as well as computational musicology can help move both fields forward. Some line here about if we really want to understand what is happening we need to know about causal factors going on here and have experimental manipulation and things like making models of the whole thing or talk about what Judea Pearl thinks about the ability to do some sort of causal modeling

with diagrams. Great to rely on some sort of anecdoatal evidence, but if we are going to put things on the line with our education then we need to be able to make some sort of falsifiable claims about what we are doing. Can only do that through the lens of science.

1.1   Claims about need to join the worlds of theory and pedagogy

- (Butler, 1997)
- (Klonoski, 2000) - perceptual hierarchy, not enough info from aural skills training
- (Karpinski, 2000) - "There is indeed a gap between the disciples of music cognition and aural skills training", GK says that one of his goals is to bridge that gap, and he does.

1.2   Chapter Overview

In this first chapter, I introduce the process of melodic dictation and discuss factors that would presumably could play a role in taking melodic dictation. The chapter introduces both a theoretical backgorund and rationale for using method form both computational musicology and congitive psychology in order ot answr quesitona bout how individuals learn melodies. I argue that tools for understanding this best because as we currently understand it, I see us operating in a Kuhnian normal science where much can be learned by just using the tools in front of us. This chapter will clearly outline the factors hypothesized to contribute to an individual's abilit to learn melodies, incorporating both individual and musical parameters. The chapter ends with a discussion some of the philosophical/theoretical problems with attempting to measure thigns like this (is it just a party trick?) and establishes that I will be taking a more polymorphic view of musicianship in order to answer this question.

The second chapter of my dissertation focuses on the history and current state of aural skills pedagogy.

Tracing back its origins to the practical need to teach musical skills back with Guido d'Arezzo, I compare and contrast the different methodological approaches that have been used, along with their goals.

The third chapter discusses previous work that examines individual factors thought to contribute to one's ability to perform an aural skills task, and it will discuss results from an experiment contributing to a discussion of how individual differences could contribute to how a person learns melodies.

Turning away from individual differences and focusing on musical features, in the fourth chapter I plan to discuss how music researchers can use tools from computational musicology as predictive features of melodies. Inspired by work from computational linguistics and information theory, recent work in computational musicology has developed software capable of abstracting features thought to be important to learning melodies, such as note density and 'tonalness' (Müllensiefen, 2009). Talk a bit about how this has been also looked at before in the music education community.

While these features have been used in large scale, exploratory studies, work in this chapter will discuss how these features could be used in controlled, experimental studies as a stand-in for the intuition many music pedagogues have when determining difficulty of a melody in a classroom setting.

In my fifth chapter, I introduce a novel corpus of over 600 digitized melodies encoded in a queryable format. This dataset will also serve as a valuable resource for future researchers in music, psychology, and the digital humanities. This chapter begins with a discussion of the history of corpus studies, noting their origin outside of music, their current state in music, and their limitations. This chapter, encapsulating the encoding process, the sampling criteria, and the situation of corpus methodologies within the broader research area, will go over summary data and also talk about how it could be used to generate hypotheses for future experiemnts (n-gram stuff based on patterns) .

Lastly, in the final chapter, I will synthesize the previous research in a series of melodic dictation experiments. Stimuli for the experiments are selected based on the abstracted features of the melodies and are manipulated

as independent variables based on the previous theoretical literature. I then model responses from the experiments using both individual factors and musical features in order to predict how well an individual performs in behavioral tasks similar to some of my previously published research (Baker & Müllensiefen, 2017). Here I also note important caveats in scoring melodic dictation, referencing some other of my own work on using metrics, such as edit distance (Baker & Shanahan, 2018), to discuss similarities between the correct answer and an individual's attempts at dictation. Results from the final chapter will be discussed with reference to how findings are applicable to pedagoges in aural skills settings. Recommendations will be made building on current conceptual frameworks (Karpinski, 2000).

Chapter 2

Theoretical Background and Rationale

2.1 What is melodic dictation? and Why?

Melodic dictation is the process in which an individual hears a melody, retains it in memory, and then uses their knowledge of Western musical notation to recreate the mental image of the melody on paper in a limited time frame. For many, becoming proficient at this task is at the core of developing one's aural skills (Karpinski, 1990). For over a century, music pedagogues have valued melodic dictation[1] which is evident from the fact that most aural skills texts with content devoted to honing one's listening skills have sections on melodic dictation (Karpinski, 2000). Additionally, any school accredited by the National Association of Schools of Music in North America requires students to learn this skill (Nat, 2018, §VIII.6.B.2.A).

Yet despite this tradition and ubiquity, the rationales as to *why* it is important for students to learn this ability often comes from some sort of appeal to tradition or underwhelming anecdotal evidence. The argument tends to go that time spent learning to take melodic dictation results in increases in near transfer abilities after an individual acquires a certain degree of proficiency learning to take melodic dictation. Rationales given for why students should learn melodic dictation has even been described by Gary Karpinski as being based on "comparatively vague aphorisms about mental relationships and intelligent listening" (Karpinski, 1990, p.192), thus leaving the evidence for the argument for learning to take melodic dictation not being well supported.

Some researchers have taken a more skeptical stance and asserted that the rationale for why we teach melodic dictation deserves more critique. For example, Klonoski in writing about aural skills education aptly questions "What specific deficiency is revealed with an incorrect response in melodic dictation settings?" (Klonoski, 2006). Earlier researchers like Potter, in their own publications, have noted how they have been baffled that many musicians do not actually keep up with their melodic dictation abilities after the class ends (Potter, 1990), but presumably go on to have successful and fulfilling musical lives. Additionally, suggesting that people who can hear music and then are unable to write it down, thus are unable to think *in* music (Karpinski, 2000), seems somewhat exclusionary to musical cultures that do not depend on any sort of written notation.

Though despite this skepticism towards the topic, melodic dictation remains at the forefront of many aural skills classrooms. The act of becoming better at this skill may or may not lead to large in increases in

---

[1]In his highly influential book *Aural Skills Acquisition: The Development of Listening, Reading, and Performing Skills in College-Level Musicians*, Karpinski (2000) documents this sentiment in music pedagogy circles by highlighting poetic adages from Romantic composer Robert Schumann in the mid 19th century through 21st century music educator Charles Elliott in the opening of his book, thus providing concrete examples of the belief that improving one's aural skills, or *ear*, is a highly sought after advanced skill.

far transfer of ability, but used as a pedagogical tool, teaching students to take melodic dictation brings with it concepts that have been deemed relevant to the core of undergraduate music training. While there has not been extensive research on melodic dictation research in recent years– in fact Paney (2016) notes that since 2000, only four studies were published that directly examined melodic dictation– this skill set sits on the border between literature on learning, melodic perception, memory, and music theory pedagogy. Understanding and modeling exactly how melodic dictation works remains as a untapped watershed of knowledge for the field of music theory, music education, and music perception and is deserving of much more attention.

In this chapter I examine literature both directly and indirectly related to melodic dictation by first reviewing the prominent four step model put forth by Karpinski in order to establish and describe what melodic dictation is. After describing his model, I then critique what this model lacks and clarify what is missing by providing a taxonomy of parameters that presumably would contribute to an individual's ability to take melodic dictation. Using this taxonomy, I then review relevant literature and assert that the next steps forward in understanding how melodic dictation works come from examing the process both experimentally and computationally. It has been nearly two decades since *Aural Skills Acquistion* was first published as the first major step to finally build a bridge between the field of music cognition and music theory pedagogy (Butler, 1997; Karpinski, 2000; Klonoski, 2000) and as with all public works, they need to be maintained?[2]

2.1.1   Describing Melodic Dictation

Much of the foundational theoretical work on the topic of melodic dictation comes from Gary Karpinski. Summarized most recently in his *Aural Skills Acquisition* (Karpinski, 2000)– though first presented in an earlier article (Karpinski, 1990)– Karpinski proposes a four-step model of melodic dictation.[3]

The four steps of Karpinski's model include

1. Hearing
2. Short Term Melodic Memory
3. Musical Understanding
4. Notation

and occur as a looping process depicted in Figure 1. The model is discussed extensively in both this original article (Karpinski, 1990) and throughout the third chapter in his book (Karpinski, 2000).

[FULL SCAN HERE EVENTUALLY OF FIGURE 1]

2.1.1.0.1   FIGURE 1 STAND IN

1. Hear (along with Portion B)
2. Remember (Portion A only)
3. Understand

- Temporal
    - Pulse
    - Meter
    - Rhythmic proportions
- Pitch
    - Tonic
    - Scale degree of starting pitch
    - Scale degree of subsequent pitches
        * Stepwise Groups
        * Each skip treated as new starting pitch

---

[2]Yes I know this is an awful metaphor and I will change it eventually

[3]This four stage process sythesizes earlier research where in Karpinski 1990 he notes two other models of melodic dictation, one from Rogers where he says there are 2 processes, another from Thomas who says there are 15

4. Notate

- What has been heard, remembered, understood

Karpinski's **hearing** stage involves the initial perceptions of the sound at the psychoacoustical level and the listener's *attention* to the incoming musical information. If the listener is not actively engaging in the task because of factors such as "boredom, lack of discipline, test anxiety, attention deficit disorder, or any number of other causes" then any further processes later down the model will be detrimentally effected. Karpinski notes that these types of interferences are normally "beyond the traditional jurisdiction of aural skills instruction", but I will later argue that the concept of willful attention, when re-conceptualized as working memory, may actually play a larger role in the melodic dictation process than is claimed here.

The **short-term melodic memory** stage in his process is where musical material is held in active memory. From Figure 1 it appears that the stage is not conceptualized as an active process where something like active rehearsal would occur, but rather just consists soley of passive mental representation. Though Karpinski does not posit any sort of active process in the short term melodic memory stage, he does suggest there are two separate memory encoding mechanisms, one for contour, and one for pitch. He arrives at these two mechanisms by using both empirical qualitative interview evidece as well as noting literature from music perception that supports this claim for contour (Dowling, 1978; Dewitt and Crowder, 1986) and literature suggesting that memory for melodic material is dependant on enculturation (Oura and Hatano, 1988; Handel, 1989; Dowling, 1990). Since its publication in 2000, this area of research has expanded with other reearchers also demonstrating the effects of musical acculturation via exposure (Eerola et al., 2009; Stevens, 2012; Pearce and Wiggins, 2012a).

In describing the short term melodic memory stage, Karpinksi also details two processes that he believes to be nesscary for this part of melodic dictation: extractive listening and chunking. Noting that there is probably some sort of capacity limit to the perception of musical material, citing Miller (1956), Karpinski explains how each strategy might be used. Extractive listening is the process in which someone dictating the melody will selectivly remember only a small part of the melody in order to lessen the load on memory. Chunking is the process in which smaller musical elements can be fused together in order to expand how much information can be actively held in memory and manipulated. The concept of chunking is very helpful as a pedagogical tool, but as detailed below, is a complicated concept to pin down how it works.

After some musical material is extracted, then represented in memory, the next step in the process is **musical understanding**. At this point in the dictation the dictator needs to take the extracted musical material that is represented in memory and the use their music theoretic knowledge in order to comprehend any sort of hierarchical relationships between notes, common rhythmic groupings, or any sorts of tonal functions. This is the point in the process where solimization of either or both pitch and rhythm, and musical material might be understood in terms of relative pitch. In the model solimization takes place later, but it is worth questioning if it is possible to dissacociate relative pitch relations from the qualia of the tones themselves (Arthur, 2018). For Karpinski, the more quickly what is represented in musical memory can be understood, the more quickly it can then be tranlated at the final step of notation.

Notation, the final step of the dictation loop, requires that the individual taking the notation have sufficient knowledge of Western musical notation so that they are able to translate their musical understanding into written notation. This last step is ripe for errors and has proved problematic for researchers attempting to study dictation (Taylor and Pembrook, 1983; Klonoski, 2006). It is also worth highlighting is that it is difficult to notate musical material if the individual who is dictating does not have the requisite musical category and knowledge for the sounds. Lack of this knowledge will limit an individual's ability to translate what is in their short term melodic memory into notation, even if it is perfectly represented in memory!

The final parts of the chapter, Karpinski notes that other factors like **tempo**, the **length and number of playings**, and the **duration between playings** will also play a role in determining how an individual will perform on a melodic dictation. While this framework can help illuminate this cognitive process and help pedagogues understand how to best help their students, presumably there are many more factors that contribute to this process. The model as it stands is not detailed enough for explanatory purposes and lacks in two areas that would need to be expanded if this model were to be explored experimentally and

Figure 2.1: Melodies of Equal Length

Figure 2.2: Melodies of Equal Length

computationally.

First, having a single model for melodic dictation assumes that all individuals are likely to engage in this sequential ordering of events. This could in fact be the case[4], but there is research from music perception (Goldman et al., 2018) and other areas of memory psychology such as work on expert chess players (Lane and Chang, 2018) that suggests that as individuals gain more expertise, their processing and categorization of information changes. Additionally, different individuals will most likely have different experiences dictating melodies based on their own past listening experience, an area that Karpinski refers to when citing literature on musical enculturation based on statistical exposure. The model does not have any flexibility in terms of individual differences.

Second, the model presumes the same sequence of events for every melody. As a general heurstic for communicating the process, this process is generalizable, but intuition would suggest that treating all melodies the same is not going to lead to having a robust model. For example, on page 103, Karpinski suggest that two listenings should be adequate for a listener with few to no chunking skills to listen to be able to dictate a melody of twelve to twenty notes. This process might generalize to many tonal melodies, but presumably different strategies in recognition would be involved in dictating the two melodies of equal length shown in Figure 2 and 3.

Presumably different people with different levels of abilities will perform differently on different melodies and while helpful as a pedagogical tool, this one size-fits-all approach to melodic dictation is not robust. This agnosticism for both variability for melodic and individual differences serves as a stepping off point for this study. In order to have a more through understanding of melodic dictation, there needs to be a model that is able to accomodate the exhaustive differences at both the individual and musical levels. Additionally, the model should be able to be operationalized so that it can be explored in both experimental and computational settings. By explicitly stating variables thought to contribute and noting how melodic dictation works, it will give the community a better sense of the melodic dictation process, which will then enable a more

---

[4]And in his Figure 3.1 he does caption it as an *idealized* dictation process
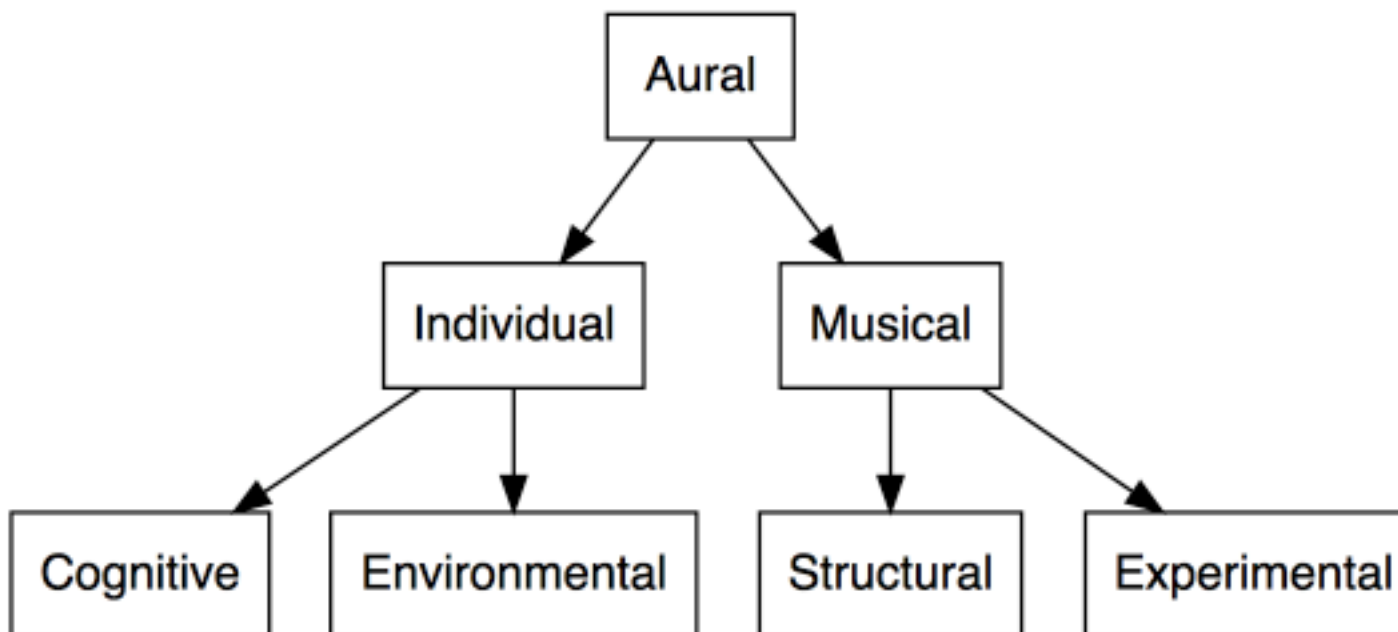
Figure 2.3: Taxonomy of Factors Contributing to Aural Skills

through understanding of melodic perception and subsequently allow for better teaching practices in aural skills classrooms.

At this point, it is worth stepping back and noting that the sheer amount of variables at play here is cumbersome and almost haphazard. In order to better understand and organize factors thought to contribute to this process, it would advantagous for future research to taxonomize the multitude of features thought to contribute to melodic dictation. In doing this, it wil allow for a clearer picture of what factors might contribute and what literatures to explore in order to learn more about them.

The taxonomy that I propose appears in Figure 4 and bifurcates the possible factors thought to affect an individual's ability to take melodic dictation into two categories: **individual** parameters and **musical** parameters. Each of these two categories can then be split again into **cognitive** and **environmental** parameters as well as **structural** and **experimental factors** respectively. Below I expand on what these categories entail, then explore each in depth.

The **individual** parameters split broadly into **cognitive factors**, or factors of people that are relatively consistent with people over time and could be understood as largely being governed by *nature*. The other category of this division consist of factors that change with training and exposure and could be understand as largely being governed by *nurture*. This second set of parameters are the **environmental** factors. These

categories are not deterministic, nor exclusive, and almost inevitably interact with one another.[5]

For example, it would be possible to imagine an individual with higher cognitive ability, the oppertunity to have a high degree of training early on in their musical career, and personality traits that are associated with higher learning aptitudes. This individual's musical perception abilities might be markedly different than someone with lower cognitive abilities, no opertunity for individualized training, come from a lower socio-economic status, and not have a general inclination to even take music lessons. This variability at the individual level might then lead to differences in their ability to take melodic dictation.

Complementing the individual differences, there would also be differences at the **musical** level which in turn divides into two categories. On one hand exists the **structural aspects** of the melody itself. These are aspects of the melody that would remain invariant when written down on a score. Parameters in this category would include features generated by the interval structure of the pitches over time that allow the melody to be categorically distinct from other melodies. These structural features are then complimented by the **experimental features** which are emergent properties of the structrual relation of the pitches over time based on performance pratice choices. Examples of these parameters would include, key, tempo, note density, timbral qualities, and the amount of times a melody is played during a melodic dictation or emergent properties like a melody's tonalness as computed through various metrics. This division is not an exhaustive, categorical divide. One could imagine exceptions to these rules where a melody is tranformed to the minor key, ornamented, and then played with extenive rubato and experienced as a phenomenologically similiar experience.

Given all of these parameters that could contribute to the melodic dictation process, the remainder of this chapter will exploring literature using this taxonomy as a guide. The chapter concludes with a reflection on operationalizing each of these factors and problems that can arise in modeling and reminds the reader about the dangers of statistical reification. These are important to note since from an empirical standpoint, both the task as well as the process of melodic dictation as depicted by Karpinksi resemble something that could be operationalized as both an experiment, as well as a computational model and if understood this way will be subjectd to the same types of critique.

2.2 Individual Factors

2.2.1 Cognitive

Research from cognitive psychology suggests that individuals differ in their perceptual and cognitive abilities in ways that are both stable throughout a lifetime and are not easily influened by short term traning. When investigated on a large scale, these abilities– such as general intelligence or working memory capacity– predict a wealth of human behavior on a large scale ranging from longevity, annual income, ability to deal with stressful life events, and even the onset of Alzheimer's disease (Ritchie, 2015; Unsworth et al., 2005). Given the strength and generality of these predictors, it is worth investigating the extent that these abilities might contribute when investigating any modeling of melodic dictation. It is imporant to understand the degree to which these cognitive factors might influence aural skills abilities in order to ensure that the types of assesments that are given in music schools validly measure abilities that individuals have the ability to improve upon. If it is the case that much of the variance in a student's aural skills grades can be attributed to something the student has little control over, this would call for a serious upheaval of the current model of aural skills teaching and assesment.

Recently there has been a surge of interest in this area [6] which could be attributed to the fact that educators are picking up on the fact that cognitive abilities are powerful predictors and need to be understood since they inevitably will play a role in pedagogical settings.

---

[5]Could footnote this interaction and talk about how AP people generally need to have genetic predisposition, fixed pitch instrumetn in house, and start at early age CITATION FOR THAT?

[6]DO I CITE THINGS HERE LIKE THAT WORK OF NANCY ROGERS, LEIGH VAN HANDLE, THE UTAH GUY, GARY KARPINSKIS ICMPC, THE FORM AT SMPC

Before diving into a discussion regarding differences in cognitive ability, I should note that sometimes ideas regarding differences in cognitive ability been hostily received (citation against people talking about IQ) and for good reasons. Research in this area can and has been taken advantage to further specious ideologies, but often arguments that assert meaningful differences in cognitive abilities between groups are founded on statistical misunderstandings and have been debunked in other literature (Gould, 1996). Considering that, it then becomes very difficult to maintain a scientific commitment to the theory of evolution (Darwin, 1859) and not expect variation in all aspects of human behavior, with cognition falling within that umbrella. Even given this statement, measuring a theoretical construct such as an aspect of cognition deserves to be examined since the ability to validly and reliably measure an individual's cognitive ability is a fundmental assumption of this study.

2.2.2 Measuring Intelligence

Attempting to measure and quantify aspects of cognition go back over a century. Even before concepts of intelligence were posited by Charles Spearman and his conception of *g* (Spearman, 1904), scientists were interested in establishing links between an individual's mental capacities and some sort of physical manifestation. The origins of this area of research have been critiqued on the basis that the early work implicitly tended to validate preconceptualized beliefs on the superiority of certain groups of peoples and used methodlogies that today would be considered risible.

- For example, BROCA thought he could get at intelligence by measuring skulls AND MORTON
- Or Spitzka who post hoc measured eminance and brain size page 127

While not immediatly relevant to current thinking in cognitive psychology, work from both Broca and XXX was continued by the American herediterian school of IQ (page 187 in Gould) and the early research done by Alfred Binet on IQ took inspiration from Broca. This lineage of ideas has often been used to tarnish systematic investigations into differences in cognitive ability, which from their outset were to initially funded by the French governement to identify children struggling in the classroom so that they could be given special attention.

Bient was the initial developer of the idea of an intelligence quotient or IQ[7] and provided one of the first ways to attempt to quantify a theoretical concept that was not capable of being manifested in the physical world. It was also around the same time that researchers like Cyril Burt and Charles Spearmean began developing their new theories of intelligence founded on the reification of factor analysis. In developing a battry of tests whose performance on one subtest could often reliably predict performance on another– a manifestation referred to as the postive manifold– Spearman and Burt put forth a separate conception of intelligence based on the ability to solve problems without any sort of background information and referred to this ability a *g* for general intellgience.

Though seemingly unrelated to the current state of thinking about cognitive abilities, Binet's and Spearman's ideologies about what intelligence is and how to measure it still represent two of the larger schools on cognitive ability. On one hand their idea that cognitive abilities are based upon a steady growth of incoming information that someone is able to manipulate once they retrieve from long term memor; on the other hand there is a school of thought that there is some sort of measureable construct, *g* that aids in the process of solving problems that do not depend on any sort of contextuxal information. Conceptualizing cognitive ability as these two different constructs inevitably leads to different types of measurements and subsequently what these constructs are then able to predict in terms of human behavior.

Without detailing entire histories of both lines of thought, Binet's conceptulization manifested into an argument for general crystallized intelligence or *Gc*, or the ability to solve problems based on previously acquired skills. Spearman and Burt's ideas about *g* school reflect a belief that individuals have some sort of latent cognitive ability to draw on to perform mental tasks. The cognitive psychology literature has noted that *g* often shares a statistically equivalent relationship to idea conceptualized as general fluid intelligence *Gf*, or the ability to solve problems in novel situations Cattell, 1971; Horn, 1994). This distinction between *Gf*

---

[7]divide mental age by chronological age then multiply by 100

and *Gc* is different than that of *g*, but again it should be noted that *Gf* and *g* share a statistically identical relationship Matzke, Dolan, and Molenaar (2010). These conceptions of intelligenec and cognitive ability aslo differ from more current theories that sythesize these previous areas of research (Kovacs and Conway, 2016).

Even though both of these constructs are powerful predictors on a large scale and do predict things like educational success, income, and even life expectancy (Ritchie, 2015)– even when other variables like socioeconomic status are held constant. Yet despite this, only conceptualizing cognitive abilities in terms of intelligence does not fully explain the diversity of human cognition.

Another large area in the field of cognitive psychology is the area of working memory capacity. In addition to concepts of intelligence, be it *Gf* or *Gc*, the working memory capacity literature also is directly relevant to work on melodic dictation for reasons discussed below.

2.2.3 Working Memory Capacity

Working memory is one of the most investigated concepts in the cognitive psychology literature. According to Nelson Cowan, the term working memory generally refers to

> the relativly small amount of information that one can hold in mind, attend to, or, technically speaking, maintain in a rapidly accesible state at one time. The term working is mean to indicate that mental work requires the use of such informaiton. (p.1) (Cowan, 2005)

The term, like most concepts in science, does not have an exact definition, nor does it have a definitive method of measurement. While there is no universally recognized first use of the term, researchers began to postulate that there was some sort of system that mediated incoming sensory information with the world with the information in long term storage using modular models of memory in the mid-twentieth century. Summarized in (Cowan, 2005), one of the first modal models of memory was proposed by (Broadbent, 1958) and later expanded by (Atkinson and Shiffrin, 1968). As seen in FIGURE X, both models here posit incoming information that is then put into some sort of limited capacity store. These modal models were then expanded on by Baddeley and Hitch (Baddeley and Hitch, 1974) in their 1974 chapter with the name *Working Memory*, where they proposed a system with an central executive module that was able to carry out active maintenance and rehearsal of information that could be stored in either a phonological store for sounds or a visual sketchpad for images.

Later revisions of their model also incorporated an episodic buffer (Baddeley, 2000) where the modules were explicitly depicted as being able to interface with long term memory in the rehearsal processes. The model has even been expanded upon by other researchers throughout its lifetime. The most relevant to this study is by (Berz, 1995), who postulated adding a musical rehearsal loop to the already established phonological loop and visual spatial sketchpad. While Berz is most likely correct in asserting that the nature of storing and processing musical information is different to that of words or pictures and there has been experimental evidence to suggest this (Williamson et al., 2010) that has been interpreted in favor of multiple loops (Wöllner and Halpern, 2016) , it does introduce the theoretical problem of multiple stores which has been addressed by other researchers.

In addressing the problem of explicitly stating which rehearsal loops do and do not exist, Nelson Cowan proposed a separate model (Cowan, 1988, 2005) dubbed the Embedded Process Model which do not claim the existence of any domain specific module (e.g. positing a phonological loop, visual spatial sketchpad) but is rather based on an exhaustive model that did away with the problem of asserting specific buffers for new types of information.

In Cowan's own words comparing his model from that of Baddeley:

> The aim was to see if the description of the processing structure could be exhaustive, even if not complete, in detail. By analogy, consider two descriptions of a house that has not been explored completely. Perhaps it has only been examined from the outside. Baddeley's (1986) approach to modeling can be compared with hypothesizing that there is a kitchen, a bathroom,

Modal Model (after Broadbent, 1958)

| Unlimited Sensory Store | Limited-Capacity Store | Long-Term Store |

Working-Memory Model (after Baddeley, 2000)

Central Executive

Passive Buffers:

| Phonological Loop | Episodic Buffer | Visuospatial Sketchpad |

Long-Term Memory (Types Related to Buffers)

Embedded-Processes Model (after Cowan, 1988)

Central Executive

Brief Sensory Afterimages ~250 msed

a

b

Activated Sensory and Categorical Features from Long-Term Memory
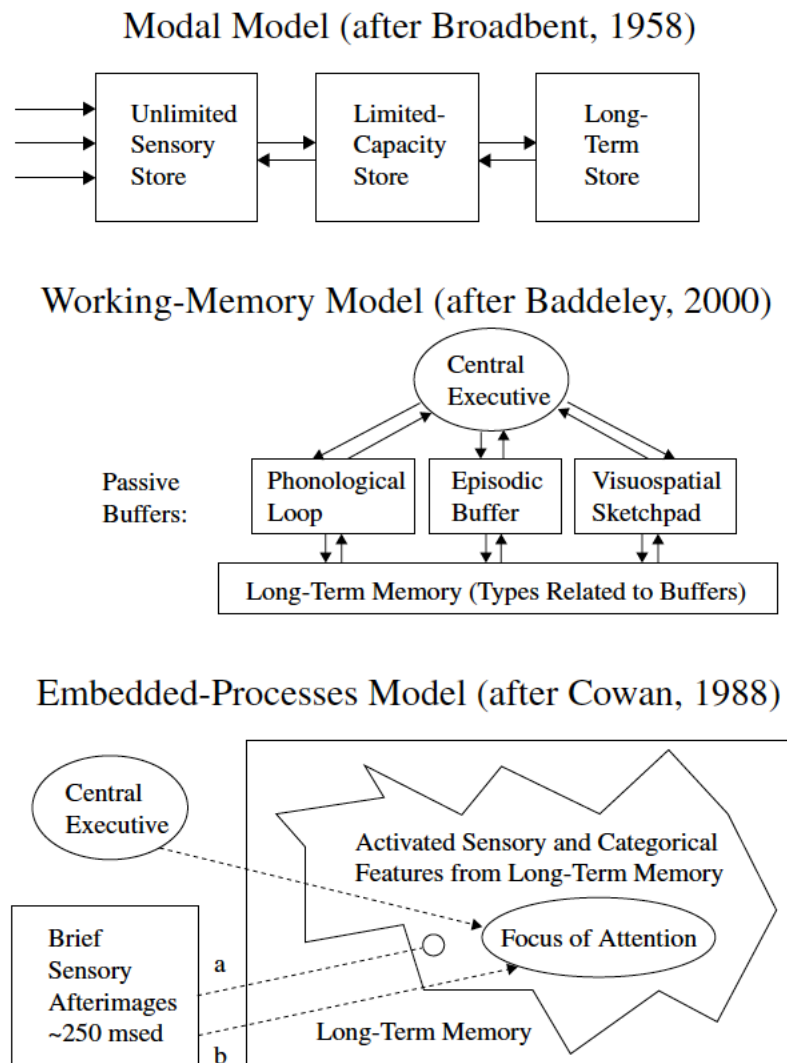
Focus of Attention

Long-Term Memory

Figure 2.4: Schematics of Models of Working Memory taken from Cowan, 2005

two equal-size square bedrooms, and a living room. This is not a bad guess, but it does not rule out the possibllity that there actually are extra bedrooms or bathroom, that the bedroom space is apportioned into two rooms very different in size, or that other rooms exist in the house. Cowan's (1988) apporach, on the other hand, can be compared with hypothesizing that the house includes food preparating quarters, sleeping quarters, batroom/toilet quarters, and other living quarters. It is meant to be exhaustive in that nothing was left out, even though it is noncommital on the details of some of the rooms. p.42 Cowan, 2005.

The system is depicted in the bottom tier of TABLE X, and conceptualizes the limited amount of information that is readily available as being in the focus of attention, with activated sensory and categorical features of what is in the focus of attention to be accessible nearby. Moving further from the locus of attention is long term memory, whose content can be turned to by using the central executive to access non-immediately available information. In contrast to the modular approaches, Cowan's framework does not require the researchers to specify exactly how and where each the incoming information is being stored which makes it advantageous for studying complex stimuli such as music and melodies.

In addition to having multiple frameworks for studying working memory capacity, there is also the problem of limits to the working memory system, often referred to as the working memory capacity. Most popularized by Miller in his famous (Miller, 1956) speech turned article, Miller suggests out of jest that the number 7 might be worth investigating, which has been used as a point of reference for many researchers since then. It is worth nothing that Miller has since gone on record as noting that using 7 (plus or minus 2) was a rhetorical device used to string together his speech (Miller, 1989). Nevertheless, while the number seven is most likely a red herring, it did inspire a large amount of research on capacity limits. In the decades since the number 7 has been reduced to about 4 (Cowan, 2010) and research around capacity limits has been investigated using a variety of novel tasks, most noteable the complex span task CITATION.
When used as predictors in both higher and lower cognitive tasks, measures of working memory capacity predict performance well and additionally tend to be stable across a lifetime (Unsworth et al., 2005).

Given its predictive strength as well as its direct similarity to tasks of melodic dictation, a in depth look at the literature is warranted.
Clearly an individual's ability to take in sensory information, maintain it in memory, actively carry out other tasks (like notating said melody) are almost identical to tasks of working memory capacity. Before venturing onward from this striking parallel, tasks investigating working memory capacity differ from melodic dictation tasks in a few key ways. The first is that musical information is always sequential: a melodic dictation task would never require the student to recall the pitches back in scrambled orders. Serial order recall is an important characteristic in the scoring and analyzing of working memory tasks (Conway et al., 2005), but musical tones do not appear in random order and are normally in discernable chunks as discussed by Karpinski(Karpinski, 2000). The use of chunks is pervasive in much of the memory literature, but often is used as more of a heuristic to help explain that information in the environment and why it is often grouped together. Of the problems with chunking, most are related to music and have relevance to melodic dictation. Below I review the problems with chunking noted by Cowan (Cowan, 2005), and any pertitant music psychology literature.

1. *Chunks may have a hierarchical organization.* Tonal music has historically been understood to be hierarchical (Krumhansl, 2001; Meyer, 1956; Schenker, 1935) with the study for memory for tones being confounded by some pitches being undestood by their relation to structurally more important tones.

2. *The working memory load may be reduced as working memory shifts between levels in hierarchy.* If an individual understands a chunk to be something such as a major triad, the load on working memory would be less since it that information could be understood as a singular chunk.

3. *Chunks may be the endpoints of a continuum of associations.* Pairing a group of tones together that might be functionally anomolous like…

4. *Chunks may include asymmetrical information.* More tonal possibilities are possible from a stable note like tonic or dominant, whereas in a tonal context, a raised scale degree $\#\hat{4}$ when understood in a functional context would be taken as having stricter transitional probabilities ($\#\hat{4} \rightarrow \hat{5}$).

5. *There may be a complex network of associations.* If a set of pitches sounds like a similar set of pitches

from long term memory , the information coming in can not be understood as being separate units of working memory.

6. *Chunks may increase in size rapidly over time.* Three tones that are seemingly unrelated when incoming like E4, G5, C5 might enter sensory perception as three different tones, but then be fused together when they are understood as one chunk– a first inverstion major triad.

7. *Information in working memory may benefit from rapid storage in long term memory.* Given the amount of patterns that an individual learns and can understand, as soon as something is fused, it could be encoded in long term memory, especially if there is a salient feature in the incoming melodic information such as the immediate recognition of a mode or cadence.

The points by Cowan are important to acknowledge in that it it not possible to directly lift work and paradigms from working memory capcacity to work in music perception. That said, the enormous amount of theoretical frameworks put forward by the working memory liteature when understood in conjunction with theories in music psychology such as implicit statistical learning (Saffran et al., 1999) can provide for new, fruitul theories. Past reserachers have noted the strength and predictive abilities of literature from the working memory capacity as aiding research in music perception. In ending his article positing a musical memory loop to be annexed to the Baddley and Hitch modular model of working memory, Berz (Berz, 1995) captures the power of this concept in the last sentence of his article and warns future reserachers that

> Individual differences portrayed in some music aptitude tests may [sic] represent not talent or musical intelligence but ability, reflecting differences in working memory capacity. p. 362

Berz's assertion has not been exhausivly tested since first published in 1995, but the subject of music, memory, and cognitive abilities has been the focus of research of both psychologists and musicologists alike. Below I survey literature bordering on both music, as well as cognitive abiltity.

2.2.3.1 Working Memory Capacity and Music

Of the papers in the music science literature that specifically investigates working memory, each uses different measures, though but all tend to converge on two general findings. The first is that there are some sort of enhanced memory capabilities in individuals with musical training. The second is that working memory capacity, however it is measured, often plays a significant role in musical tasks. Evidence for the first point appears most convincingly in a recent meta analyses by Talamini and colleagues (Talamini et al., 2017) who demonstrated via three separate meta-analyses that musicians outperform their non-musical counterparts on tasks dealing with long-term memory, short-term memory, as well as working memory. The authors also noted that the effects were the strongest in working memory tasks where the stimuli were tonal, which again suggests an advantage of exposure and understanding of the hierarchical organization of musical materials. In this meta-analyss and others investigating music and cognitive ability, it is important to be reminded that the direction of causality still from these studies cannot be determined using these statistical methodologies. While it might seem that musical training tends to lead to these increases, it is also possible that higher functioning individuals will self select into musical activities. Even if there is no seletion bias in engaging with musical activity it also remains a possiblity that of the people that do engage with musical activity, the higher functioning individuals will be less likely to quit over a lifetime.

In terms of musical performance abilities, working memory capacity has also been shown to be a significant predictor. Kopiez and Lee suggested that working memory capacity should to contribute to sight reading tasks based on research where they found measures of working memory capacity, as measured by a matrix span task, to significantly correlated with many of their measures hypothesized to be related to sight reading ability in pianists at lower difficulty grading (Kopiez and Lee, 2006, 2008).

Following up on this work on sight reading, Meinz and Hambrick (Meinz and Hambrick, 2010) found that working memory capacity, as measured by an operation span task, a reading span task, rotation span task, and a matrix span task was able to predict a small amount of variance $R^2 = .074(0.067)$ above and beyond that of deliberate practice alone $R^2 = .451(.441)$ in a sight-reading task. More recently, two studies looking at specific sub groups of musicians have shown working memory capacity to significantly contribute to models of performances on musical tasks related to novel stimuli. (Wöllner and Halpern, 2016) found that although no

differences were found between pianists and conductors in measures of working memory capacity as measured via a set of span tasks, conductors showed superior performance in their attention flexibility. Following up on this line of research (Nichols et al., 2018) used the same battery of working memory tasks and found that jazz musicians excelled over their classically trained counterparts in a task which required them to hear notes and reproduce them on the piano. The authors also noted that of their working memory battery, based on standard operation span methods (Engle, 2002), that the auditory dictation condition scored surprisingly low and further research might consider further work on dictation abilities. Additionally (Colley et al., 2018) found that working memory capacity, as measured by a backwards digit span and operation span, to be successful predictors in a tapping task requiring sensory motor prediction abilities. As mentioned above, each of these tasks where working memory was a significant predictor of performance occured where the task involved active enagement with novel musical material.

The growing evidence in this field suggests that the advantage of working memory capacity to be greatest in both musically trained people, dealing with novel information, using tonal materials. Since all three of these factors are related to melodic dictation, it would seem sensible to continue to include these measures in tasks of musical perception and continue Berz's assertion that research in music perception could inadvertently be picking up on individual differences in working memory abilities.

2.2.4   General Intelligence

As discussed above, the idea of IQ or intelligence has a long history that is bothgood and bad. When used as a predictor in models it often serves to predict traits that society values like longevity and general income so given its ability to predict in more domain general settings, surveying literature where it applies to musical activity is a worthy task. Below I use the term intelligence as a catch all term to avoid the historical context of IQ and specify where availible which measure was actually used. Before surveying the literature here it is also worth noting that reserach on music and intelligence is not as developed as some of the larger studies looking at intelligence which provides problems for both establishing causal directionality, as well as controlling for other factors like self theories of ability, socioeconomic status, and personality (Müllensiefen et al., 2015).

2.2.4.1   Papers that suggest GF plays a role

As reviewed in Schellenberg (2017), both children and adults who engage in musical activity tend to score higher on general measurses of intelligence than their non-musical peers (Gibson, Folley and Park, 2009; Hille et al., 2011; Schellenberg, 2011a; Schellenberg and Mankarious, 2012) with the duration of training sharing a relationship with the extent of the increases in IQ (Degé, Kubicek and Schwarzer, 2011a; Degé, Wehrum, Stark and Schwarzer, 2015; Corrigall and Schellenberg, 2015; Corrigall, Schellenberg and Misura, 2013; Schellenberg, 2006). Though many of these studies are correlational, they also have made attempts to control for confounding variables like socio-economic status and parental involvment in out of school activities. (Corrigall et al., 2013; Degé et al., 2011a; Schellenberg, 2006, 2011a, 2011b; Schellenberg and Mankarious, 2012). Schellenberg notes the problem of smaller sample sizes in his review (Corrigall and Trainor, 2011; Parbery-Clark et al., 2011; Strait, Parbery-Clark, Hittner and Kraus, 2012) in that studies that are typically smaller do not reach statistical significance. Schellenberg also references evidence that when professional musicians are matched with non-musicians from the general population there do not seem to be these associations (CITE). Schellenberg's review suggests the current state of the literature points to support the hypothesis that higher functioning kids that take music lessons and they tend to stay in lessons longer. Additionally, Schellenberg remains skeptical of any sorts of causal factors regarding increases in IQ (e.g., François et al., 2013; Moreno et al., 2009) noting methodlogical problems like how short exposure times were in studies claiming increases in effects or researchers who not holding pre-exisiting cognitive abilitis constant (Mehr, Schachner, Katz and Spelke, 2013).

- (Corrigall et al., 2013a)
- (Corrigall et al., 2013b)

- (Swaminathan et al., 2017)

2.2.5 Environmental

Standing in contrast to factors that individuals do not have a much control over such as the size of their working memory capacity or factors related to their general fluid intelligence, most of the factors we believe contribute to someone's ability to take melodic dictation have to deal with factors related to training and the environment. In fact, one of the tacit assumptions of getting a music degree revolves around the belief that with deliberate and attentive practice, that an individual is able to move from novice to expertise in their chosen domain. The idea that time invested results in beneficial returns is probably best exemplified by work produced by Ericsson et al. (1993) that suggests that performance at more elite levels results from deliberate practice. Below I review literature that supports this argument, since it's no doubt that someone has to engage in something to be good at it.

2.2.6 Musical Training

- Papers that suggest practicing makes you better?

- It almost seems redundant to review literature in support of music practice leading to better results.

- List of those papers here

2.2.7 Aural Training

In additon to individuals differeng in their general musical abilities– however they are defined– individuals also differ in their abilities at the level of their aural skills. Though not heavily researched in the past few decades (Furby, 2016), there has been specific research looking at explaining how people do in aural skills examinatons. Harrison et al. (1994) examined the effect of aural skills training on undergraduate students by creating a latent variable model investigating musical aptitude, academic ability, musical expertise, and motivation to study music in a sample of 142 undergraduate students and claimed to be able to explain 73% of the variance in aural skills abilities using the variables measured. Work from Colin Wright's dissertation found that … (Wright, 2016)

These are things that people have suggested people trying to do : As noted in Furby (2016), researchers in the past have suggested a variety of techniques for improving their abilities in melodic dictation by isolating rhythm and melody [Banton (1995); Bland (1984); Root (1931); WILSON], listening attentively to the melody before writing (Banton, 1995), recognizing patterns (Banton, 1995; Bland, 1984; Root, 1931) and silently vocalizing while dictating (Klonoski, 2006).

2.2.8 Sight Singing

Often described as the other side of the same coin of melodic dictation, sight singing is an area of music pedagogy research that has had sparse attention paid to it given its prevelance in school of music curricula. Recently (Fournier et al., 2017) catalouged and categorized 14 different sub categories into four larger main cateogories while also providing commenatary on some of the current state of aural skills. Of the four large categories, they group them into reading mechanisms, sight singing, readings skills acquiskiton, and learnign support.

FIX HERE

The authors note a line of research that has documented that university students are often unprepared to sight-read single lines of music (Asmus, 2004; Davidson, Scripp & Welsh, 1988; Fournier, 2015; Thompson, 2004; Vujović & Bogunović, 2012) even though it is, like dictation, thought of as a means for deeper musical

understanding. (DeBellis, 2005; Karpinski, 2000; Ottman, 1956; Rogers, 2004; Scripp, 1995; Scripp & Davidson, 1994) The authors of Fournier et. al also note that sight-reading has been an active area of research due to the often reported relationship that performance on sight reading often predicts several studies have shown links between academic success in sight-singing and predictors such as entrance tests (Harrison, 1987, 1990, 1991; Ottman, 1956; Rodeheaver, 1972; Schleuter, 1983), academic ability (Chadwick, 1933; Harrison, 1990, 1991; Harrison, Asmus, & Serpe, 1994; Rodeheaver,1972), and musical experience (Brown, 2001; Dean, 1937; Furby, 2008; Harrison, 1990, 1991; Harrison et al., 1994; Thostenson, 1967).

Taken as a whole, the research tends to suggesting that learning to be a fluid and compotent sight reader helps musicians hone their skills by bootstrapping other musical skills since the skills needed for sightreading touch on many of the skills used in musical performance like pattern matching and listening for small changes in intonnation. Each of these individual factors contributes in a small and significant way and additionally will interact with the other half of the taxonomy: musical parameters.

2.3 Musical Parameters

Transitioning to the other half of the taxonomy on figure X, the other main source of variation on any study looking at melodic perception, and consequently studies of melodic dictation, is the effect of the melody itself. I find it safe to assume that not all melodies are equally difficult to dictate and assert that variance in the difficulty the melody can partitioned between both **structural** and **experimental** aspects of a melody. As noted above, there is not a strict deliniation between these two categories since once could imagine drastic maniuplations in experimental parameters in order to result in a phenomenologically different experience of melody. Questions of transformations of melodies and musical similarity fall have been addressed in ohter research (Cambouropoulos, 2009; Wiggins, 2007) but are beyond the scope of this study.

2.3.1 Structural

The notion that the music as represented by a score is able to provide insights towards understanding aspects about music is not new to music theory and anlaysis. Heinrich Schenker argued for an undertanding of tonal music (Schenker, 1935) that asserted hierarchical relationships of notes on the musical surface in the early 20th century and has since been expanded upon by theorists over the past century SALTZER, SCHAKTER, ROTHSTEIN. Leonard Meyer in his *Emotion and Meaning in Music* (Meyer, 1956) continued some of these lines of thought and was the first who put forward the idea that tethered the struture that earlier theorists wrote about and suggested that in addition to this struture being fundamental to the perception of the piece that there were also responsible for some aspects of the emotion and meaning listeners found in the piece.

Meyer's work since inspired a line of reserach investigating the perception of the structural aspects of music could be understood with the work of Eugene Narmor (Narmour, 1990, 1992), Glenn Schellenberg (Schellenberg, 1997), Elizabeth Hellmuth Margulis (Margulis, 2005), David Huron (Huron, 2006) and have inpsired computational, machine learning approaches to expectational frameworks in with work by Marcus Pearce (Pearce and Wiggins, 2012a).

- WHAT IS THE POINT OF ALL THIS RESERACH SETNECE

These general models of melodic perception tend suggest that Meyer was correct in his assertion that computational methodologies could be used to better understand question of melodic perception and strucuture and that there are links between the structure of the music and its perception.[8]

Turning to studies examining melodic dictation with a focus on musical structure, the first study to examine it extensively was Ortmann in 1933 (Ortmann, 1933). Ortamnn used a series of twenty five-note melodies in order to examine the effects of repetition, pitch direction, conjunct-disjunct motion (contour), interval size, order, and chord structure, all of which he deems to be the *determinants* of an individual's ability to

---

[8]This whole section sucks

dictate melodic material. Though Ortmann did not use any statistical methods to model his data, he did assert that each of his determinants contributed to an individual's ability to dictate musical material. This work was extended by (Taylor and Pembrook, 1983) which additionally incoporated using musical skill as a predictor and additionally found evidence that these factors contributed to individual dictation abilities in a sample of 122 undergraduate students.

Although the literature is generally sparce compared to other areas of music cognition, literature exploring the effects of structural characteristics on memory does exist. Long found that length, tonal structure, contour, and individual traits all contribute to performance on melodic dictation examinations and found that structure and tonalness to have significant, albeit small predictive powers in modeling (Long, 1977). One problem with studies such as (Long, 1977) is that they sometimes would make conspicuous methodloogical decisions such as elimnating individuals who were bad singers for the example. Not only does this reduce the spectrum of ability levels (assuming that singing ability correlates with dictation ability, a finding since which has been established (Norris, 2003)), but is additionally flawed in that it is at odds both with the intuition that an individual's singing ability cannot be taken as a direct representation of their mental image of the melody and is probably more related to the ability to have motor control over the vocal tract (Pfordresher and Brown, 2007).

Other researchers have also put forward other paramters thought to contribute like tempo (Hofstetter, 1981) tonality (Dowling, 1978) (Long, 1977) (Pembrook, 1986) (Oura and Hatano, 1988) interval motion (Ortmann, 1933; Pembrook, 1986) length of melody (Long, 1977; Pembrook, 1986) number of presentations (Hofstetter, 1981) [(Pembrook, 1986)] context of presentation (Schellenberg and Moore, 1985) listener experience (Long, 1977; Oura and Hatano, 1988) (Schellenberg and Moore, 1985; Taylor and Pembrook, 1983) familarity of style (Schellenberg and Moore, 1985)

Pembrook (1986) provides an extensive detailing of a systematic study to melodic dictation where they used tonality, melody length, and type of motion as variables in their experiment. They additionally also restricted their experimental melodies to those that were singable. The authors found all three variables to be significant predictors with tonality explaining 13% of the variance, length explaining 3% of the variance and type of motion explaining 1% of the variance. The paper also claims that people on average can hear and remember 10-16 notes, which is worth commenting that these 10–16 notes are dependent on the experimental context of the melodies played with the quarter note set to 90 beats per minute.

Given the lack of consistent methodlgies in adminstration and scoring of these experiments it becomes difficult to find ways to generalize basic findings like expected effect sizes– especially when the original materials and data have not been recorded– but there is often interesting theoretical insights to be gleaned. For example (Oura, 1991) used a sample of eight people to suggest that when taking melodic dictation, individuals use a system of pattern matching that interfaces with their long term memory in order to complete dictation tasks. While this paper does not bring with it exhaustive evidence supporting this claim, the idea is explored in detail in Chapter 6 the idea of pattern matching is used in conjunction with Cowan's embeded process model of working memory.

More recently the music education community has also began to do research around melodic dictation using both qualitative and quantitiatve methdologies. (Paney and Buonviri, 2014) interviewed high school teachers on methods they used to teach melodic dictation and (Gillespie, 2001) has done work on investigating methods as to best score melodic dictation. Other work by (Pembrook and Riggins, 1990) surveyed various methodlgies used by instructurs in aural skills settings. Some of these studies consider aural skills as a totality like (Norris, 2003) who provided quantiative evidence to suggest most aural skills pedagoge's intuition that there is some sort of relationship between melodic dictation and sight singing. Looking at the notorious subset of students with absolute pitch (AP), (Dooley and Deutsch, 2010) provided evidence demonstrated that students with AP tend to outperform their non-AP colleagues in tests of dication.

[SOME OF THESE ARE CLEARLY EXPERIMENTAL AND NEED TO BE PUT THERE]

Continuing exploring the pedagogical liteature, Naton Buonviri and colleagues have also made melodic dictation a central focus of some recent papers. Paney and Buonviri (2014) interviewed high school teachers on methods that they used to teach melodic dictation. Buonviri (2015a) interviewed six sophomore music

majors to find sucessful strategies that students enaged with when completing melodic dications. Paney (2016) reported beneficial effects to direct student's attention and guide them through melodic dictation exercises suggesting that some sort of mental organizaton of the dication process is helpful. Buonviri and Paney (2015) found that having students sing a preparatory singing pattern after hearing the target melody, essentially a distractor task, hindered performance on melodic dictation. Buonviri (2015b)... Buonviri (2015a) found no effects of test presentation format (visual versus aural-visual) using a melodic memory paradigm. Buonviri (2017) reported no significant advatage to listening strategies while partaking in a melodic dictation test.

2.3.1.1   Recent Computational Musicology Work papers and findings

Using symbolic features of the melodies themselves is not a novel approach as noted in the above literature attempting to predict performance on melodic dictations. Much of this work pre-dates recent advances in computational musicology such as the advent of technology like David Huron's Humdrum (Huron, 1994) and Michael Cutberth's Music21 (Cuthbert and Ariza, 2010) which now allows music researchers to systematically digitize symbolic musical material. In addition to creating accesible frameworks for encoding, the computational power availble exponentially exceeds that of what was availble in the 20th century and has opened up new possibilities in the computational modeling of music.

While I reserve a longer discussion on the histories of computational musicology for the fourth chapter, relevant to this study is the additional ways it is now possible to abstract features from symbolic melodies beyond what was capable in studies such as Ortmann (1933) and Taylor and Pembrook (1983).

An abstracted feature of a melody is an emergent property of the melody that results from performing some sort of calculation on the melody? This type of feature abstraction is in contrast to much of the work done in the field of music information retrival which often relies on the recorded audio for feature abstraction and is addressed under Experimental features LINK THAT IN!. Abstracted symbolic features of melodies can largely be conceptualized as being **static** or **dynamic**. The above papers tended to use more simplisitc methods of figuring out parameters such as counting the notes by hand but with the advent of new encoding systems and more powerful computing power it is now possible to take on much more rigerious computational analyses.

2.3.1.2   Static Views of Computational Features/ FANTASTIC

Static features of melodies work by summerizing some aspect of the melody as if it were to be experience in suspended animation. Using static features helps quantify something that might be intuitive about a melody or piece of encoded music. For example, something like David Huron's contour class used in a study investigating melodic arches (Huron, 1996) using the Essen Folk Song Collection (Schaffrath, 1995) can only be understood as a feature of the melody itself once the melody has been sounded and is recalled would be a static feature of a melody. Other examples include a melody's global note density, normalized pairwise variability index (CITATION), and a melody's tonalness as calculated by one of the various key profile algorithms (KRUMHANSL,ALBRECT AND SHANAHAN) These measures are useful when describing melodies and are predictive of various behavioral phenomena as detailed below, but at this point it has not been well established to what degree these summary features can be directly related to aspects of human behavior.

The quintessential and most comprehensive toolbox example of this is Daniel Müllensiefen's Feature ANalysis Technology Accesing STatistics (In a Corpus) or FANTASTIC (Mullensiefen, 2009). FANTASTIC is software that is capable summerizing musical material at for monophonic melodies. In additon to computing 37 features such as contour variation, tonalnesss, note density, note length, and measures inspired by computational linguistics (THAT BOOK OF GERAINT), FANTASTIC also calculates m-types (melodic-rhyhmic motives) that are based on the frequency distributions of melodic segements found genres of music. This is inspired by fact that repeition is key structure of music (Huron, 2006)

Work using the FANTASTIC toolboox has been sucessful in predicting court case decisions (Müllensiefen and Pendzich, 2009), predicting chart sucesses of songs on the Beatles' *Revolver* (Kopiez and Mullensiefen,

2011), memory for old and new melodies in signal detection experiments (Müllensiefen and Halpern, 2014), memory for earworms (Jakubowski et al., 2017; Williamson and Müllensiefen, 2012), memorability of pop music hook (Balen et al., 2015). In experimental studies, FANTASTIC has also been used to determine item difficulty (Baker and Müllensiefen, 2017; Harrison et al., 2016) and has even been the basis of the development of a computer assistted platform for studying memory for melodies (Rainsford et al., 2018).

2.3.1.3   Dynamic

In addition to using summary based features on melodies, it is also possible to model the perception of musical materials by using a dynamic approach that is dependent on the unfolding of musical material. First explored in CONKLIN, and then first published as a dynamic model of expectation in his doctoral dissertation, Marcus Pearce's Informaton Dynamics Of Melody IDyOM models musical expectancy using various information theoretic concepts inspired by Claude Shannon (SHANNON). The model takes an unsupervised machine learning approach and calculates the information content of the amount of DECLARED n-grams in the corpus. As a model exploring expection for melody IDyOM has has been applied to a variety of settings LIST THEM HERE. The domain general application of IDyOM has given creedence to Meyer's assertion that the enculturation of musical styles stems from statistical exposure to melodies and be somewhat refelective of the cogitive processes used in muscical perception. IDyOM has also been recently extended to look at expectation in multi-part chorales (SAUVE WORK) and expectations of harmony (HARRISON WORK).

(Pearce and Wiggins, 2012b; Pearce, 2018)

The advantage of using a dynamic approach is that it theoretically reflects real time perception of music with the structural characeristics of the music mapping on to real human behavior.

2.3.2   Experimental

- Advantage of vocal melodies paper
- MIR stuff
- Stuff out of education stuff that is also experimental

2.4   Modeling and Polymorphism of Ability

Given the current state of cognitive psychology and psychometrics, as well as recent advances in computational musicology, the possibilities for now operationalizing and then modeling aspects of melodic dictation are as advanced as they ever have been. Given that we can now assign numbers to basically every factor that is thought to contribute to this process from concepts of musicianship, to features of a melody, to the variable size in an individual's working memory capacity all of these thigns can be put into some sort of model. While this will bring the community closer to formally modeling all of this and lead to a clearer understanding, before going ahead and doing this it is worth pointing out that many of the concepts discussed above are highly complex concepts like musicianship and tonalness and rest on lots of assumptions. Musicianship, for example, or any measure of musical training is not something that can be measured directly such as a person's height or weight, but has to be inferred based on the logical assumptions of the person doing the measurments. So while the rest of this study will rely on this, it is important to note that people shouldn't confuse abstracted concepts with real things.

The most illustrative example of this comes from a study by HARRISON ET AL who created a latent variable model of aural skills that was able to predict 74% of the variance in aural skills performance. This latent trait that the authors created may be helpful in explaining the patterns of covariance in data, but this would be to reifiy a statistical abstraction as an ontologically true idea. This idea has been discussed before critqiuing ideas such as g (Gould, 1996; Kovacs and Conway, 2016) and has recently been the subject of critique in music psychology OUR GOLDMSI PAPER.

The same arguments put forward in this literature also are relevant here. In order to have a complete, causal model of *how* melodic dictation works, it is important to understand melodic dictation as a set of musical abilities that are related to other musical abilities, though may not be related. This idea is not new even in music psychology, the past two decades have seen calls for a more polymorphic definition of musical ability (Levitin, 2012; Peretz and Coltheart, 2003) which in its modeling will require more concrete ways of defining how it works than just correlating variable together that are helpful at prediction without saying exactly how that process happens.

2.5   Conclusions

In this chapter I first described what is melodic dictation using Karpinki's verbal model, noted what the things were that were missing from this model as a stepping off point, then went on to suggest a taxonomy of these based on what already has done. I suggest there are both individual as well as musical features that need to be understood in order to have a comprensive understanding of melodic dictation. Of the two sets of features, individual features can be either cognitive or environmental and musical features can be either structural or experimental. This taxonomy does not consist of exclusive categories and certainly permits interactions between any and all of the levels.

It would be impossible given the scope of this study to effectively quantify each and every factor and how it interacts at every level, but the degree to which there is the most literature and you can get the most bang for you buck seems like the obvious stepping off point. Rest of this dissertation will systehsize these areas and put forth novel research contributing to the modeling and subsequent understanding of melodic dictation. Understanding melodic dictation will help with both understanding melodic perception and help our pedagogy.

2.5.1   Add In

- (Hansen et al., 2013)
- Dowling 1991 (Dowling, 1991)

Chapter 3

Individual Differences

3.1   Why care about cognitive abilities

3.1.1   General intelligence and WMC

3.1.2   Defining of terms

3.2   Have established that cognitive abilities contribute to musical task (for journal article langauge repeat)

3.2.1   General Fluid Intelligence, WMC, Training as uni of polymorphic

3.3   Remind the nature of a musical dictation type task (hear, loop, executive decision)

3.3.1   This is WMC task, gf has problems (Although high level link with gf, problematic, WMC models at level of process of md)

3.3.1.1   Berz 1994 noticed it first

3.3.1.2   Williamson Baddely Hitch suggest maybe musical loop

3.3.1.3   Even Cowan labs wonder how different (Li Cowan Saults)

3.4   WMC has been misused in music education, theory, pedagogy, aural literature and deserves attention

3.4.1   Problems with chunking

3.4.1.1   Mistake with Miller 1956, he did not mean 7 items

3.4.1.2   Broadbent 1956 more of why its more like 3-4

3.4.2   Problems with using capacity limit literature

3.4.2.1   See Cowan 2005 page 80

3.4.2.2   Musical order is always serial effects

Chapter 4

Computation Chapter

OTHER PEOPLE WHO HAVE DONE THIS

- look into Wiggins et al., 1993, for history of representation

Folk music

- Bartok 1936?

- Bartok and Lord 1951

- Lomax 1977

- Steinbeck 1982

- Jesser 1992

- Sagrillo 1999

- GET AND READ PAT SAVAGE ARTICLE

Popular Music

- Moor 2006
- Kramarz 2006
- Furnes 2006
- Riedemann ????

Computational Musicology

- Eerola eta al 2007 and 2007
- McCay 2005
- Huron 2006
- Frieler 2008
- JAZZOMAT PROJECRT OUTPUT

4.1 Humans like patterns and are very good at picking them up

4.1.1 We learn things implicitly

4.1.2 We can represent that implicit knowledge with a corpus

4.2 Pre-Musical Corpora

4.2.1 Information Theory

4.2.2 Computational Linguistics as front runner

4.3 Musical Corpora

4.3.1 History of Musical Corpora

4.3.1.1 Fun old computational music papers

4.3.1.2 Corpora that are often used

4.3.1.3 Static vs Dynamic models of feature abstraction (daniel slides?)

4.3.2 FANTASTIC

4.3.2.1 static

4.3.2.2 ML approach gets it right

4.3.2.3 simple to understand

4.3.2.4 Can abstract features be percieved?

4.3.2.4.0.1 Note density

4.3.2.4.0.2 Contour variation

4.3.2.4.0.3 Tonalness

4.3.2.4.0.4 weird computational measures

4.3.3 IDyOM as representation of musical materials

4.3.3.1 n-gram models

4.3.3.2 mirrors human behavior

4.3.3.2.0.1 melody

4.3.3.2.0.2 harmony

4.4 So What?

4.4.0.1 Other research (Chapt 3) suggest need to move beyond cognitive measures

4.4.0.2 Can operationalize item level items contextually with a corpus

4.4.0.3 IF features are real, they should effect dictation (Chater 6)

4.4.0.4 Not only important for one off, but then would be incorporated into computational learning models (Chapter 6)

4.4.0.5 We need new materials

Chapter 5

Hello, Corpus

5.1 Brief review of Chapter 4 on corpus (Language to reflect journal submission)

5.1.1 Corpus outside of music

5.1.2 Corpus in Music

5.1.3 The point is that it implicitly represents humand knowledge

5.1.4 IDyOM 1

5.1.5 IDyOM 2

5.1.6 IDyOM 3

5.1.7 Huron suggestions that starts of melodies relate to mental rotaiton

5.1.8 Other Huron claims

5.2 Note problem with using corpus is making corpus

5.2.1 Many are used on Essen

5.2.2 Brinkman says Essen Sucks

5.2.3 If going to make generlizable claims, need to always have new data

5.3 Solem duty to encode and report on corpus

5.3.1 Justin London Article on what makes it into a corpsu

5.3.2 Though I just encoded the whole thing because in my heart of hearts I'm a Bayesian

5.4 The Corpus

5.4.1 History of Sight Singign books

Chapter 6

Experiments

6.1 Rationale

6.1.1 Have done all this and have not actually talked about dictation yet

6.1.2 Clearly many factors contribte to this whole thing and need to be taken into a model

6.1.3 Dictation is basically a within subjects design Experiment

6.1.3.1 Get very ecological and dirty and run it

- Paney 2016 had 30 second timing

6.1.4 Factors

6.1.4.1 Cognitive

6.1.4.1.1 WMC

6.1.4.1.2 GF

6.1.4.2 Training

6.1.4.2.1 Goldsmiths MSI

6.1.4.3 Musical

6.1.4.3.1 FANTASTIC

6.1.4.3.2 IDyOM

6.1.4.4   Investigate melodies with this context and set scoring

6.1.4.5   Mirror design to see if effects of melody are there

6.2   Experiments

6.2.1   Experiment I

6.2.1.1   Participants

6.2.1.2   Procedure

6.2.1.3   Materials

6.2.1.4   Scoring

6.2.1.5   Results

6.2.1.6   Modeling

6.2.1.7   Discussion

6.2.2   Experiment II

6.2.2.1   Participants (New)

6.2.2.2   Procedure (Same)

6.2.2.3   Materials (Swapped but controlled)

6.2.2.4   Scoring (Same)

6.2.2.5   Results

6.2.2.6   Modeling (same)

6.2.3   General Discussion

6.2.3.1   What happened

6.2.3.2   Assumption of all of this is that many things are happening linearly in combination with each other

6.2.3.3   Additionally the mixed effects framework works better with more data?

6.2.3.4   Also how we score it is going to mess wiht the DVs

6.2.4   Really what is needed is Computational Model

Chapter 7

Computational Model

7.1   What Is A Model And Why do you need a model

In his 2007 article *Models of music similarity* (Wiggins, 2007) Geraint Wiggins makes a distinction between *descriptive* and *explanatory* models when describing models that describe human behavior. Descriptive models assert what will happen in response to an event. For example, as noted in _____. While the _____ is assumed to drive the rise in _____, merely stating that there is an established relationship between one variable and the other says nothing about the inner workings of this process. An explanatory model on the other hand not only describes what will happen, but additionally notes why and how this process occurs. For example, much of the work musical expectation demontrates that as exposure to a musical style for an individual increases, so does that persons's ability to to predict specific events with a given musical texture. Not only does it predict more accurate responses, but many of these models like HURON, MARGULIS, PEARCE derrive their underlying theory about this behavior from the brain's ability to implicitly track statistical regularities in musical perception SAFFRAN. The *how* derrives from the brain's ability to track statistical regularities musical information in the environment and the *why* derrives from evolutionary demands in that organisms that are able to make more predictions about their environment are more likely to survive and pass on their genes HURON 2006. Wiggins notes that although there can be both explanatory and descriptive theories, depending on the level of abstraction, a theory may be explanatory at one level, yet descriptive at another. Using the mind-brain dichotomy, he notes that something like a theory of implict expectations like PEARCE could be explanatory at the level of behavior as noted above, but says nothing about what is happening at the neuronal level. Both descriptive and explanatory theories are needed in that descriptive theories are used to test explanatory theories, and by strinng together different layers of abstraction, we can arrive at a better understanding of how the world works.

Returning to melodic dictation, under Wiggins' framework, the Karpinki model of melodic dictation (Karpinski, 2000, 1990) would qualify as a decriptive model. The model says what happens over the time course of a melodic dictation– specificying four discrete stages discussed in earlire chapters– but does not explicitly state *how* or *why* this process happens. In order to have a more complete understanding of melodic dictation, an explanatory model is needed.

In this chapter I introduce an explanatory model of melodic dictation. The model is inspired by work from both computational musicology, as well as cognitive psychology and uses a uniform memory mechanism inspired by Cowan's Embedded Process Model of working memory to explain the process. In addition to quantifying each step, the model has points in it where parameters could be tweaked in order to accomodate for individual differences, while still relying on a domain general process. The model consists of three main parts, each with its own set of parameters that

1. Prior knowledge
2. Selective Attention

3. Transcription

Inspired by Baysian computational modeling, the *Prior Knowledge* part reflects the previous knowledge an individual brings to the melodic dictation. The selective memory– somewhat akin to Karpinksi's extractive listening[1]– segments incoming musical information by using the window of attention/limits of working memory capacity as a sensory bottleneck to constrit the size of musical chunk that an individual were to trascribe. Once the focus of attention, the Transcription function recusrively pattern matches against the prior knowledge's corpus of information in order to find a match of explicitly known musical information. This process refects, but does not actually mirror the exact process used in melodic dictation, yet seems to be phenomenologically similar to the decision making process used when attempting notate novel melodies. Based on both the prior knowledge and individual differences of the indiviual, the model will scale in ability, with the general retrevial mechanisms in place. The exact details of the assumptions, free parameters, and complete formula of the model are discussed below.

- Name it here if you want

It is important to note that this model is not perfect in that it deterministically mirrors the exact cognitive process someone would use to take melodic dictation. It instead quantifies each step of the process and provides a detailed account each step of the process in a fully realized and implmentable process. The benefits of creating a computational model of melodic dictation are numerous. First, as stated above, a computational model acts as an explanatory model that details the inner working of a process. Second, by quantifying each step of the process, a computational model forces the researcher to specify key things that could be blown past. In this case, we have to deal with the representation of musical material, defintions of working memory, as well as what does it mean for something to have been learned. Thirdly, in creating a computational model you provide theories with a concrete version that can be more easier faslfied. The quintessential example of the importance of this comes from an anecodote in evolutionary psychology claiming that in heterosexual populations, men engage in more casual sexual encounters than women due to their promiscous proclivities. This theory has appeared in textbooks, but as noted by AUTHOR, while might be good as a verbal model in that it plays to our preconcived notions about male sexuality, implemented as a computational model, is mathematically impossilble. Fourthly, creating a computational model rules defines the universe of a possible outcomes that your theory posits. As shown in work done on the seemingly simple Baddely and Hitch model of working memory, AUTHOR demonstrated that with XXX parameter values, the YEAR model of working memory actually allows for 15X different permuatations of the model. With such vast possibilites, it becomes much harder to speak about_the_ model as a single entity. Finally, and most importantly from a pedagogical framework, by detailing the exact steps of the process, it provides a theory of melodic dictation in which each element of the process is able to be discussed and gives exact framework for which it can be criqtiqued. Creating a computational model is the next step in furthering research on aural skills pedagogy.

7.2   The Model

7.2.1   Model Representational Assumptions

In order to write a computer program that mirrors the melodic dictation process, how the mind "percieves" and "thinks" about musical information must be defined *a priori*. Before delving into questions of representation, this model assumes that the musical surface JACKENDOFF as represented by the notes via Western musical notation are salient and can be percieved as distnict perceptual phenomena. Although there is work that suggests that different cultures and levels of experience might not categorize melodic information universally MCDERMOTT, other work suggests that things like experiencing pitches as discrete, categorical phenomena is categorized as a statistical human universal SAVAGE. For the purposes of this model in that it models something specific to a very WEIRD CITATION demographic, I think it is safe.

---

[1]I use different terms to avoid confusion when comparing models later

Knowing that it is melodic information or melodic data that needs to be represented, the question then becomes what are the best ways in which to represent it. Much work has been done on this history of musical representation WIGGINS ET AL 1993 has been exhastive, even though it has only been on for about 30 years but regardless of what exact representation is chosen, it generally can be concived of a symbol string Clifford et al., 2005. How a string is represented carries with it important assumptions about a listern's percepetion and also a teacher's pedagogy. Take for example the many ways in which someone could represent the ever popular Twinkle, Twinkle Little Star as in Figure 1.

- TWINNLE IMAGE HERE in C and F#

Representing the string of symbolys as in Figure 2A using THIS NOTATION SYSTEM would suggest that listeners perception and cognition of the notes mirrors that of fixed *do* solfedge. The only symbolys in the string, or tokens, represented carry with it the name of the note along with it's absolute range. When compared to string 2B– the melody transposed a tritone up– a computer would find no symbol overlap between the two strings and might declare the two as completely dissimilar. Additionaly, neither string contains with it any sort of rhythmic information. Reproduced Figure 2C and Figure 2D, we see the same two strings now with additionally temporal information.

- FIGURE 2 DIFFERENT TYPES OF REPRESENTATION
- C4 C4 G4 G4 A4 A4 G4
- F#4 F#4 C#5 C#5 D#6 D#6 C#6

As stated above, this mapping assumes fixed pitch representation of musical information. This is at odds to intuition in that most humans would categorize the informaiton in strings 2A and 2B as containing the same musical content.

In order to capture that of human perception and cognition, one might want to represent this string of musical informationt to the computer as a list of directional intevals (Math citation) which brings with it interval information not present (although maybe impliclty imposed!) in the previous examples.

- C U +P5 U +M2 U -M2

This then captures what might be considered phenomena closer to human perception, but again is devoid of any sort of scale degree functionality or qualia that might be involved in listener's perception. Including this assumption the string and durational values as it might be a Humdrum spine might be represented as:

- DO DO ^SOL SOL ^LA LA >SOL (with rhythm)

Each string represnts what might be considered the same perceptual pheneomena, but when quantified, one system needs to be explicitly chosen as how information is represented will change model output. For example if it is assumed that an individual's prior musical information can only be encoded as the first *fixed do* example, then each of the 12 possible ascending major fifths would have to be learned as separate phenomena to be represnted in as prior knowledge in this model. This point also brings up the importance of noting that many of these systems could be at play here and just because one representation might not make sense in one implemntation of a model does not make it null and void for other applications in pedagogical settings.

For the purposeses of this model and further examples I will use the absolute interval representation of pitches here, with further directions in this research addressing how it would be possible to implment things like the scale degree and *fixed do* model. Comparisons of this model output will also contribute to conversations regarding pedagogy in that if one form of data represenation mirrors human behavior than the others, it would provide more than anecodotal evidence in support of the pedagogy of one system over another. How the model represents musical information is the first important parameter value that needs be chosen before running the model and this establishes the prior information.

7.3   Modeling Information Content

Having established that the models first established parameter to be decided is how strings of melodies are represented, the next important decision that has to be made is how decisions of segmentation for the second stage of attentive memory are decided. Although there has been a large amount of work on different ways to segment the musical surface JACKENDOFF DANIEL OTHERS THAT MOZART THING, rely on matching a music theorist's intution with a set of descriptive rules somewhat like the boundary formation rules in *GTTM*. While this continues a long tradition of research trying to determin formal boundaries (History of All Music Theory), since melodic dictation is an active memory process, rather than a semi-passive process of listening (tho MARIUS HAS SAID OTHERWISE/EMBODIMENT) this model needs to be able to quantify musical information that is both dependent on prior musical experience and allows a moveable boundary for selective attention so that musical information that is memory can be actively maintained, whilst carring out another cognitive process, that of notating the melody.

In order to create this metric, want to use IDyOM which uses IC (Shannon) and quantifies the IC of melodies based on a corpus of materials. Great thing about it is that IDyOM is machine learning so that can put in lots of different symbolys and get the IC. THIS IS THE WAY IT WORKS ON HIGH LEVEL WITH N-GRAMS.

Given that each grouping of n-grams will have its own IC, can then use that and cumIC function to determine. Now all you have to do is set a theshold of memory (allow for indivudal differences) and then to also allow only a certain amount into attentional memory.

Here parameter to consider is the threhold of WM which then allows for differnces at the indiviudal level.

For example, when trained against a corpus of melodies, THIS MELODY has IC (unitless thing) that is on top of notes. The advantage of operationalizing how an individual hears a melody like this is that melodies with lower information content, derrived from an understanding of having more predictiable patterns from the corpus, will allow for larger chunks to be put inside of the attentive memory buffer.

- FIGURE HERE

Note that the notes above the melody here are dependent on what is current in the prior knowledge. A corpus of prior knowledge with less melodies would lead to higher information content measures for each set of notes, while a prior knowledge that has extensive tracking of the patterns would lead to lower IC.

7.3.1   Setting Limits with Transcribe

With each note then quantified with a measure of information content, it then becomes possible to set a limit on the maximum amount of information that the individual would be able to hold in memory as defined by a an attentional threshold. A higher threshold would allow for more musical material to be put in the attentional buffer, and a lower threshold would restrict the amount of information held in an attentional buffer. By putting a theshold on this value, this serves as something akin to a perceptual bottleneck basesd on the assumption that there is a capacity limit to that of working memory. Modulating this boundary will help provide insights into the degree to which melodic material can be retained between high and low working memory span individuals.

In practice, notes would enter the attentional buffer until the information content from the melody is equal to the memory threshold. At this point, the notes that are in the attentional buffer are sequestered and will be actively maintained in the attentional buffer. In theory, the maximum of the attenional buffer should not be reached since the indivudal performing the dictation would still need mental resources and attention to actively manipulate the information in the attentional buffer for the process of notating.

7.3.2 Pattern Matching

With subset of notes of the melody now represented in the attentional buffer, whether or not the melody becomes notated depends on whether or not the melody or string in the buffer can be matched with a string that is explicly known in the corpus. * Do I introduce the explict implict thing earlier? Mirroring a search pattern akin to Cowan's Embeded Process model, the individual would search across their long term memory for anything close to or resembling the string pattern in the attentional buffer.

- Image of wallpaper search here?

Using this logic, longer pattern strings n-grams would be less likley to be recalled exactly. If a pattern is not found in the initial search, one token of the n-gram would be dropped off the string and the search would happen again. This recursive process would happen until an explict long term memory match is made. Like humans taking melodic dictation, the computer would have the best luck finding patterns that fall within the largest density of a corpus of intervals distribution. Additionally, like students performing a dictation, if a student does not explicitly know an interval, or a 2-gram, the dictation would not be able to be completed. If this happens, both the model and student would have to move on to the next segement.

Upon the sucessful explict match of a string, that section of the melody would be consiered to be sucessfully dictated. The model here would register that one iteration of the funciton has been run and the chunk transcribed would then be recorded. After recording this history, the process would happen again starting at either the next note from where the model left off, the note in the entire string with the lowest information content, or n-gram left in the melody with that is most represented in the corpus. This parameter is defined before the model is run and the question of dictation re-entry certainly warrents further research and investigation.

This type of pattern search also dependent on the way that the prior knowledge is represented. In the example here, both interval and rhythm are represented with each token in the string. Since there is probably a very low likelyhood of finding an exact match for every n-gram with both pitch and rhythm, this patten search can happen again with both rhythms and pitch informaiton queried separatly. If not exact pitch-temporal matches are found and the search is run again on either the pitch or rhythmic information separatly, this would be computationally akin to Karpinki's proto-notation that he suggests students use in learning how to take melodic dictaiton (KARPINKSI CHAPTER 3). With this feature of the model incorporated, it would be possible to also generate evidence in favor or disfavor of attempting to dictate elements of the melody separately such as doing rhythm first, then pitch information or vice versa. Either way, having this as a second level default of the model seems to match with the intutions strategies that someone dictating a melody might use.

7.3.3 Loop Back IN

Upon the sucessful pattern match of a string, the attentional buffer and transcribe-search function would need too then be run again. As noted above, re-entry in the melody could be a highly subjective point of discussion. The model could either re-enter at the last note where the function sucessfully left off, the note in the melody with the lowest information content, or the n-gram most salient in the corpus. Entering at the last note not transcribed is logical from a computational standpoint, but this linear approach seems to be at odds with anecdotal experience. Entering at the note with the lowest information content seems to provide a intuitive point of re-entry in that it would then be easier to transcribe. Entering at the most represented n-gram seems to match the most with intutition in that people would want to tackle the easier tasks first, but this rests on the assumption that humans are able to reliably detect the sections of a melody that are easist to transcribe based on implicitly learned statistcal patterns. For example, some people might instead choose to go to the end of a melody after sucessful transcription of the start of the melody. This might be because this part of hte melody is most active in memory due to a recency effect, or it could be that that cadential gestures are more common in being represnted in the prior knowledge.

7.3.4   Completion

Given the recursive nature of this process, the model if all 2-grams are explicitly represented in the prior knowledge should be transcribed. If only represented using such a small chunk, the model will have to loop over the melody many times, thus indicating that the transcriber had a high degree of difficulty dictating the melody. If there is a gap in explict knowledge in the prior knowledge, only patches of the melody will be recorded and the melody will not be recorded in its entirety. An easier transcritpion will result in less iterations of the model with larger chunks. Though the current instatiation of the model does not incorporate how multiple hearingings might change how a melody is dictated, one could constrain the process to only allow a certain number of iterations to reflect this. Of course as a new melody is learned it is slowly being introduced into long term memory and could be completly be capable of being represented in long term memory without being explicitly notated at the end of a dictation with time running out and thus not possible to be completed. This of course then would be imposing some sort of experimental constraint on the process and since this is meant to be a cognitive computational model of melodic dictation, this complicates things. Future reserch could be done to optimize the choices that the model makes in order to satisfy whatever constraints are imposed and could be an interersting avenue of future resarch, but are beyond the initial goals of the model.

The model then outputs each n-gram transcribed and can be counted as a series with less attempts mapping to an easier transcription. I believe that this lines with many intutitions about the process of melodic dictaiton. It first creates a linear mapping of attempts to dictate with difficulty of the melody. It relies on a distinciton between explict and implicit statistical knowledge. It is based on the embeded processs model from working memory and attention, so is part of a larger generative model, giving it a bit more creditiblity that htis *could* be how melodic dictation works.

Having now gone through the model verbally and notating the parameters that need to be set and how things are then represnted, I will now define the model formally.

7.4   MoDiMe

- Formal Model here

7.5   Example

Having now both described the model and defined it formally, I now can run through a brief example of how the model might work in practice by showing a melody to be dictated, a the corpus of melodies that it represents, and three iterations of model that exhaust each branch of the model.

7.5.1   Example Corpus

- 12 Melodies on on page and their representation
- Also show rhtymic and pitch abstractions
- Show their density histograms as a corpus
- Target Melody
- Define WMC low for example
- Linear re-entry
- First run through take n- gram chunk, no explict pattern match, drop 1 gram, yes match, move on

- Next no exact pattern match on linear re-entry, query pitch (yes, proto) query rhythm (yes,rhythm), both transcribed

- Third iteration, full pattern match

- Not complete, but at this point have X amount of melody transcribed with this history.

- Tedious to get the exact thing, go to actual example

- CRAZY THAT YOU COULD USE THIS TO WORK BACKWARDS HOLY SHIT BAYES YOU ARE SWEET!!

7.5.2 Actual Data

- Let's now say that we have set of XYZ melodies as targets, the whole semester 1's melodies the prior knowledge, what does the model then say.
- This basically mirrors week 1 of semseter 2 of freshman year
- Plot of correlation of model predicts, correlation of actual scores
- Can re-do this now with only pitch scores (or run it with just pitch!)
- Can re-do this now iwht only rhythm socre

And here is some commentary

7.6 Future Directions

Now with a formal model of melodic dictation, what is now possible?

- Talk explicitly about representation
- Talk about what happens when
- Link to sight singing
- Testable hypotheses?
- Problems with how to define prior
- Future Directions
- How this all relates to helping teach
- aka explain the model, thus explain learning, perceptoin, process, representation, if one rep performs better than another, atomistic singing

---

7.7 Computational Cognitive Model Model (If time permits) [Whole article in itself]

COMPUTATIONAL MODEL

So thinking all about this, first reviewing the literature and then trying to run experiments to figure it all out, got kind of frustrated because even with a mixed effects model, all it is saying is that as note density goes UP so does difficult, and it interacts with tonality. Def setting myself up for a "I could have told you that" moment from my music theory colleagues. Also figured since I am doing a dissertation called modeling melodic dictation, would be good to actually make an explanatory model of it.

Hopefully by this point, have a good idea of the factors that might contribute, my task here was to then try and operationalize everything and make a computer do it.

Actually really inspired by this Lewandowsky book where they had a quip about how wit the Baddeley model it really has like 150 permutations of it and some famous example of someone saying that men have more one night stands than women, but in a heterosexual population that is mathematically impossible. So it only seemed like it made sense to see if I could write one out.

So this is how I think it would work, currently in the process of coding it.

So essentially it's a bayesian inspired model that compares a target melody, with a corpus of melodies or the prior, that are meant to represent all "known" musical material in someone's musical understanding. I got the idea from reading about Cowan's Embedded process model which does not posit the working memory structure as very distinct from that of everything in long term representation, but rather uses the central executive idea as the limited window of attention that can then be focused on things either in LTM or near categorical features of what is in the purveiw of attention.

Will first go through it on a high level, then make another pass and talk about how everything is getting operationalised. So imagine that we have a prior corpus of all the melodies that are covered for sight singing and dictaiton in the first semester of aural skills (in this case, melodies X–XX). If we let the corpus M represent all prior knowledge, where not only is the melody represented as a string of intervals, but we also get ever n gram permutation of of the melody based on the idea that the more frequent the exposure of an n-gram, the more likely it would be to be understood.

This is the first big parameter M.

We could imagine and visualize this as a grid of distributions of all possible n-grams and the strings that make them up seen here GGPLOT2. Now although any string of notes is possible, in order to mirror the limited capacity of WMC, there needs to be some sort of threshold that is set that would mirror the limits of the window of WMC.

This threshold is the next big parameter, T.

Currently conceptualizing this as being reflective of maximum of information content threshold as calculated by IDyOM.

Idea here is that as you hear more notes, more IC, the more your fixed capacity bin of melody is going to be filled up. By making it based on the IC of a melody string represented in the corpus, patterns that are more frequent are going to easier to then dictate, which aligns with intuitions on melodic dictation.

Also need to define a the explicit, implicit threshold. Explicitly known n-grams should have some sort of fixed amount that if it succedes that, you would know it. Like karpinski "knowing" phase, would also assume that every interval class would also be "known" and if not, that is OK too, will lead to people not being able to transcribe it.

We've now defined both the individual knowledge parameters and teh thrshold T of working memory capacity. These consist of WHAT DO YOU CALL THIS.

With these individual parameters established, we can calculate difficulty of the melody by running what I am going to be determining as the 'transcribe' function, which is separate from the individual parameters.

We then can introduce the target melody , t which represents the string to be dictated.

Idea here is that this melody is compared to the corpus, M, and you get the IC of each of the n-grams from the prior corpus.

Given the theshold, take the biggest n-gram that fits below the threshold to be put in the buffer. If you had a huge threshold and could take IC of whole melody or you KNEW it, would be reflective here. Could also introduce a function here that just gueses at about the threshold. Once this is filled up, let's say 4 notes, the notes are then set to the transcription buffer.

At this point index the n-gram is cross checked against the prior knowledge and look for matches of it that happen above a certain "known" threshold. If it is known, it gets "notated' and you re-enter the melody. If not, in this case the 4 gram would get recursively truncated until found a string, even 2 gram where it is

known. If it gets to a 2 gram and it's not explicitly known, basically because you don't know that interval, what Karpinksi rails againt as atomistic hearing. Also good in that people with bigger chunking will do better.

If there is not an exact match, move to split fuzzy searches (future versions)

Re-entry function is defined by finding the next part of the melody where either next most common n-gram occurs or where it left off last. This is then to reflect the two different ways of approaching tacking. Seing as some people start at front, others start at end. Could be reflective of primacy and recencey effects of memory, OR maybe it's due to fact that more prototypical n-grams in habit starts and ends of melodies….?

This process would then re-occur a few times over until either can't go anymore (intervals are not known) or till completion.

Afterwards have an idea of the path that it took, counts of those, use those as proxies of difficulty.

I think this covers everything in the decision making process and each step of the process can be automated eventually (post doc anyone?!). More importantly, it reflects ecological phenomonologcial experience of taking melodic dictaiton.

Aligns with intuitions that People who know more melodies will be better (singer phenomena) People will try for bigger chunks first, then go to smaller ones Atomistic transcription can happen but is inefficient People with higher chunking ability will do better IC is helpful proxy for this in line with IDyOM literature Relative pitch is where it's at for things like this What's deal with AP?

Chapter 8

Reference Log

8.1 To Incorporate

- (Margulis, 2005) – Margulis Model
- (Nichols et al., 2018) – Specialty jazz background helps in tasks, WMC
- (**?**) – Fix intext
- (Schumann and Klauser, 1860) – Quote about why people should do ear training
- (Smith, 1934) – Quote from K2001 about why people should do ear training
- (Long, 1977) – Musical Characteristics predict memory
- (Taylor and Pembrook, 1983) – Great citation that lots of things change memory, even structural!
- (**?**) – Long boring talk on STM, LTM
- (Oura, 1991) – Awful experimental design that says people use structual tones
- (Buonviri, 2014) – Call for experimental, suggestions as to what factors might contribute, use of deductive reasoning, qualitative
- (Buonviri, 2015b) – People need to focus right away, not establish, distractors
- (Buonviri, 2015a) – Showing people visual music does not help much.
- (Buonviri, 2017) – Listening helps with other things, no best strategy in terms of writing
- (Buonviri and Paney, 2015) – Literature to say people are bad at teaching melodic dictation and we don't know a lot about it, also interesting stuff about what solfege systems people use
- (Butler, 1997) – Call for music educators to do aural skills research, notes problem with aural skills pedagogy in lack of direction, also nice Nicholas Cook quotes on point of theory
- (Furby, 2016) – music ed study with weird stats, has references to follow up on with advantages of pitch systems and people who reccomend things for sight singing
- (Pembrook, 1986) – Effects of melodies, also how people do it. Interesting that they too effect of melodies, but talka bout things in terms of notes and not in terms of information content. Thought ot have an experiment where the n-grams that are more common are easier to write down. Lots of good charts too.
- (Paney, 2016) – It's not good if you tell people what to do when they are dictating, article has a lot of good review for dictation materials to add to the 'toRead' folder.
- (Fournier et al., 2017) – Good references that people are awful at Aural Skills, Also suggestions that people are not that great at transfer, and some stuff to suggest academic abililty is intertwined in all of this. Good reference for when starting to talk about untangling the mess that is aural skills.
- (**?**, 1995) – Add on a new module to the WMC model of baddel with music, presents some evidence for why this theoretically should be included, but actually takes examples of dictation. A lot of this article felt like things that i was reinventing…not good.
- (**?**) – Proof some other people are starting to think in terms of pedagogical schemas
- (Klonoski, 2000) – Music cognition needs to talk to aural skills more, also need to unbind theory routine with aural skills and think of things more as in a perceptual learning hierarchy

- (Klonoski, 2006) – great quotes that when people get something wrong with aural skills, what does that even mean, lack of transfer effects, article ends with ways to get better at things
- (Pembrook and Riggins, 1990) – Survey of what people in the late 1980s were doing in terms of aural skills pedagogy
- (**?**) – addresses why Gary Karpinski thinks we should teach melodic dictation
- (Potter, 1990) – dictation teacher surprised that people don't keep up their dictaiton skills quote

8.2   Chapter 3

- (Cowan, 2005) – This book will probably serve as cornerstone of chapter in terms of creating relevant literature in addition to EE course readings on WMC. Provides history of WMC models and notes how attention based model as opposed to Baddely loop might actually be better theoretical model for talking about fact that WMC could just be something related to attention if not that. Provides extensive listing on problems with chunking that are all relevant to music, but then also supports it. Shows that Miller 1956 is a generally bad citation, own author even says that in Miller 1989 (check and add) and says limit is probably about 4 (use Cowan 2001 for ctation find that). Lots of good ideas like how music is always serial recall, examples of how to model the process, great discussions on zooming out and categorical nature of music within span of WMC ideas.

- (Ockelford, 2007) – uses case of savant to argue bits of Berz WM Music Model

Bibliography

(2018). National Association of Schools of Music Handbook. Technical report, National Association of Schools of Music, Reston, Virginia.

Arthur, C. (2018). A Perceptual Study of Scale-degree Qualia in Context. *Music Perception: An Interdisciplinary Journal*, 35(3):295–314.

Atkinson, R. C. and Shiffrin, R. M. (1968). HUMAN MEMORY: A PROPOSED SYSTEM AND ITS CONTROL PROCESSES! *Human Memory*, page 54.

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

Baddeley, A. D. and Hitch, G. (1974). Working Memory. In *Psychology of Learning and Motivation*, volume 8, pages 47–89. Elsevier.

Baker, D. J. and Müllensiefen, D. (2017). Perception of Leitmotives in Richard Wagner's Der Ring des Nibelungen. *Frontiers in Psychology*, 8.

Balen, J. V., Burgoyne, J. A., and Bountouridis, D. (2015). Corpus Analyis Tools for Computational Hook Discovery. *Proceedings of International Society for Music Information Retrevial 15*.

Banton, L. (1995). The Role of Visual and Auditory Feedback during Sight-Reading of Music. *Psychology of Music*, 23(1):3–16.

Berz, W. L. (1995). Working Memory in Music: A Theoretical Model. *Music Perception: An Interdisciplinary Journal*, 12(3):353–364.

Bland, L. (1984). *Sight Singing Through Melodic Analysis: A Guide to the Study of Sight Singing and an Aid to Ear Training Instruction*. Scarecrow Press.

Broadbent, D. E. D. E. (1958). *Perception and Communication*. Pergamon.

Buonviri, N. (2015a). Effects of music notation reinforcement on aural memory for melodies. *International Journal of Music Education*, 33(4):442–450.

Buonviri, N. O. (2014). An Exploration of Undergraduate Music Majors' Melodic Dictation Strategies. *Update: Applications of Research in Music Education*, 33(1):21–30.

Buonviri, N. O. (2015b). Effects of a Preparatory Singing Pattern on Melodic Dictation Success. *Journal of Research in Music Education*, 63(1):102–113.

Buonviri, N. O. (2017). Effects of Two Listening Strategies for Melodic Dictation. *Journal of Research in Music Education*, 65(3):347–359.

Buonviri, N. O. and Paney, A. S. (2015). Melodic Dictation Instruction. *Journal of Research in Music Education*, 62(2):224–237.

Butler, D. (1997). Why the Gulf between Music Perception Research and Aural Training? *Bulletin of the Council for Research in Music Education*, (132).

Cambouropoulos, E. (2009). How similar is similar? *Musicae Scientiae*, 13(1_suppl):7–24.

Colley, I. D., Keller, P. E., and Halpern, A. R. (2018). Working memory and auditory imagery predict sensorimotor synchronisation with expressively timed music. *Quarterly Journal of Experimental Psychology*, 71(8):1781–1796.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., and Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5):769–786.

Corrigall, K. A., Schellenberg, E. G., and Misura, N. M. (2013a). Music Training, Cognition, and Personality. *Frontiers in Psychology*, 4.

Corrigall, K. A., Schellenberg, E. G., and Misura, N. M. (2013b). Music Training, Cognition, and Personality. *Frontiers in Psychology*, 4.

Cowan, N. (1988). Evolving Conceptions of Memory Storage, Selective Attention, and Their Mutual Constraints Within the Human Information-Processing System. *Psychological Bulletin*, 104(2):163–191.

Cowan, N. (2005). *Working Memory Capacity*. Working memory capacity. Psychology Press, New York, NY, US.

Cowan, N. (2010). The Magical Mystery Four: How Is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science*, 19(1):51–57.

Cuthbert, M. S. and Ariza, C. (2010). Music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data. page 7.

Darwin, C. (1859). *On the Origin of Species*. Routledge.

Dewitt, L. A. and Crowder, R. G. (1986). Recognition of Novel Melodies after Brief Delays. *Music Perception: An Interdisciplinary Journal*, 3(3):259–274.

Dooley, K. and Deutsch, D. (2010). Absolute pitch correlates with high performance on musical dictation. *The Journal of the Acoustical Society of America*, 128(2):890–893.

Dowling, W. J. (1978). Scale and Contour: Two Components of a Theory of Memory for Melodies. *Psychological Review*, 84(4):341–354.

Dowling, W. J. (1990). Expectancy and attention in melody perception. *Psychomusicology: A Journal of Research in Music Cognition*, 9(2):148–160.

Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics*, 50(4):305–313.

Eerola, T., Louhivuori, J., and Lebaka, E. (2009). Expectancy in Sami Yoiks revisited: The role of data-driven and schema-driven knowledge in the formation of melodic expectations. *Musicae Scientiae*, 13(2):231–272.

Engle, R. W. (2002). Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science*, 11(1):19–23.

Ericsson, K. A., Krampe, R. T., and Tesch-Romer, C. (1993). The Role of Deliberate Practice in the Acquisition of Expert Performance.

Fournier, G., Moreno Sala, M. T., Dubé, F., and O'Neill, S. (2017). Cognitive strategies in sight-singing: The development of an inventory for aural skills pedagogy. *Psychology of Music*, page 030573561774514.

Furby, V. J. (2016). The Effects of Peer Tutoring on the Aural Skills Performance of Undergraduate Music Majors. *Update: Applications of Research in Music Education*, 34(3):33–39.

Gillespie, J. L. (2001). Melodic Dictation Scoring Methods: An Exploratory Study. *Journal for Music Theory Pedagogy*, 15.

Goldman, A., Jackson, T., and Sajda, P. (2018). Improvisation experience predicts how musicians categorize musical structures. *Psychology of Music*, page 030573561877944.

Gould, S. J. (1996). *The Mismeasure of Man*. WW Norton & Company.

Handel, S. (1989). *Listening: An Introduction to the Perception of Auditory Events*. MIT Press, Cambridge.

Hansen, M., Wallentin, M., and Vuust, P. (2013). Working memory and musical competence of musicians and non-musicians. *Psychology of Music*, 41(6):779–793.

Harrison, C. S., Asmus, E. P., and Serpe, R. T. (1994). Effects of Musical Aptitude, Academic Ability, Music Experience, and Motivation on Aural Skills. *Journal of Research in Music Education*, 42(2):131.

Harrison, P. M., Musil, J. J., and Müllensiefen, D. (2016). Modelling Melodic Discrimination Tests: Descriptive and Explanatory Approaches. *Journal of New Music Research*, 45(3):265–280.

Hofstetter, F. T. (1981). Computer-Baesed Recognition of Perceptual Patterns and Learning Styles in Rhythmic Dictation Exercises. *Journal of Research in Music Education*, 29(4):265–277.

Huron, D. (1994). The Humdrum Toolkit: Reference Manual.

Huron, D. (1996). The Melodic Arch in Western Folk Songs. *Computing in Musicology*, 10:3–23.

Huron, D. (2006). *Sweet Anticipation*. MIT Press.

Jakubowski, K., Finkel, S., Stewart, L., and Müllensiefen, D. (2017). Dissecting an earworm: Melodic features and song popularity predict involuntary musical imagery. *Journal of Aesthetics, Creativity, and the Arts*, 11(2):112–135.

Karpinski, G. (1990). A Model for Music Perception and Its Implications in Melodic Dictation. *Journal of Music Theory Pedagogy*, 4(1):191–229.

Karpinski, G. S. (2000). *Aural Skills Acquisition: The Development of Listening, Reading, and Performing Skills in College-Level Musicians*. Oxford University Press.

Klonoski, E. (2000). A Perceptual Learning Hierarchy: An Imperative for Aural Skills Pedagogy. *College Music Symposium*, 4:168–169.

Klonoski, E. (2006). Improving Dictation as an Aural-Skills Instructional Tool. page 6.

Kopiez, R. and Lee, J. (2006). Towards a dynamic model of skills involved in sight reading music. *Music Education Research*, 8(1):97–120.

Kopiez, R. and Lee, J. (2008). Towards a general model of skills involved in sight reading music. *Music Education Research*, 10(1):41–62.

Kopiez, R. and Mullensiefen, D. (2011). Auf der Suche nach den "Popularitätsfaktoren" in den Song-Melodien des Beatles-Albums Revolver: Eine computergestützte Feature-Analyse [In search of features explaining the popularity of the tunes from the Beatles album Revolver: A computer-assisted feature analysis]. *Musik und Popularitat. Beitrage zu einer Kulturgeschichte zwischen*, pages 207–225.

Kovacs, K. and Conway, A. R. A. (2016). Process Overlap Theory: A Unified Account of the General Factor of Intelligence. *Psychological Inquiry*, 27(3):151–177.

Krumhansl, C. (2001). *Cognitive Foundations of Musical Pitch*. Oxford University Press.

Lane, D. M. and Chang, Y.-H. A. (2018). Chess knowledge predicts chess memory even after controlling for chess experience: Evidence for the role of high-level processes. *Memory & Cognition*, 46(3):337–348.

Levitin, D. J. (2012). What Does It Mean to Be Musical? *Neuron*, 73(4):633–637.

Long, P. A. (1977). Relationships Between Pitch Memory in Short Melodies and Selected Factors. *Journal of Research in Music Education*, 25(4):272–282.

Margulis, E. H. (2005). A Model of Melodic Expectation. *Music Perception: An Interdisciplinary Journal*, 22(4):663–714.

Meinz, E. J. and Hambrick, D. Z. (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):914–919.

Meyer, L. (1956). *Emotion and Meaning in Music*. University of Chicago Press, Chicago.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information.

Miller, G. A. (1989). *A History of Psychology in Autobiography*, volume VIII. Stanford University Press, Stanford, CA, l. gardner edition.

Mullensiefen, D. (2009). Fantastic: Feature ANalysis Technology Accessing STatistics (In a Corpus): Technical Report v1.5.

Müllensiefen, D. and Halpern, A. R. (2014). The Role of Features and Context in Recognition of Novel Melodies. *Music Perception: An Interdisciplinary Journal*, 31(5):418–435.

Müllensiefen, D., Harrison, P., Caprini, F., and Fancourt, A. (2015). Investigating the importance of self-theories of intelligence and musicality for students' academic and musical achievement. *Frontiers in Psychology*, 6.

Müllensiefen, D. and Pendzich, M. (2009). Court decisions on music plagiarism and the predictive value of similarity algorithms. *Musicae Scientiae*, 48:257–295.

Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. University of Chicago Press, Chicago.

Narmour, E. (1992). *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. University of Chicago Press.

Nichols, B. E., Wöllner, C., and Halpern, A. R. (2018). Score one for jazz: Working memory in jazz and classical musicians. *Psychomusicology: Music, Mind, and Brain*, 28(2):101–107.

Norris, C. (2003). The Relationship Between Sight Singing Achievement and Melodic Dictation Achievement. *Contributions to Music Education*, 30(1):39–53.

Ockelford, A. (2007). A Music Module in Working Memory? Evidence from the Performance of a Prodigious Musical Savant. *Musicae Scientiae*, 11(2_suppl):5–36.

Ortmann, O. (1933). Some tonal determinants of melodic memory. *Journal of Educational Psychology*, 24(6):454–467.

Oura, Y. (1991). Constructing a Representation of a Melody: Transforming Melodic Segments into Reduced Pitch Patterns Operated on by Modifiers. *Music Perception: An Interdisciplinary Journal*, 9(2):251–265.

Oura, Y. and Hatano, G. (1988). Memory for Melodies among Subjects Differing in Age and Experience in Music. *Psychology of Music*, 16:91–109.

Paney, A. S. (2016). The effect of directing attention on melodic dictation testing. *Psychology of Music*, 44(1):15–24.

Paney, A. S. and Buonviri, N. O. (2014). Teaching Melodic Dictation in Advanced Placement Music Theory. *Journal of Research in Music Education*, 61(4):396–414.

Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation: Enculturation: Statistical learning and prediction. *Annals of the New York Academy of Sciences*, 1423(1):378–395.

Pearce, M. T. and Wiggins, G. A. (2012a). Auditory Expectation: The Information Dynamics of Music Perception and Cognition. *Topics in Cognitive Science*, 4(4):625–652.

Pearce, M. T. and Wiggins, G. A. (2012b). Auditory Expectation: The Information Dynamics of Music Perception and Cognition. *Topics in Cognitive Science*, 4(4):625–652.

Pembrook, R. G. (1986). Interference of the Transcription Process and Other Selected Variables on Perception and Memory during Melodic Dictation. *Journal of Research in Music Education*, 34(4):238.

Pembrook, R. G. and Riggins, H. L. (1990). "Send Help!": Aural Skills Instruction in U.S. Colleges and Universities. *Journal of Music Theory Pedagogy*, 4(1):230–241.

Peretz, I. and Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, 6(7):688–691.

Pfordresher, P. Q. and Brown, S. (2007). Poor-Pitch Singing in the Absence of "Tone Deafness". *Music Perception: An Interdisciplinary Journal*, 25(2):95–115.

Potter, G. (1990). Identifying Sucessful Dictation Strategies. *Journal of Music Theory Pedagogy*, 4(1):63–72.

Rainsford, M., Palmer, M. A., and Paine, G. (2018). The MUSOS (MUsic SOftware System) Toolkit: A computer-based, open source application for testing memory for melodies. *Behavior Research Methods*, 50(2):684–702.

Ritchie, S. (2015). *Intelligence: All That Matters.* All that matters. Hodder & Stoughton.

Root, F. (1931). *Methodical Sight-Singing: Lessons in Vocal Culture.* Bryn Mawr, Pennsylvania: Theodore Presser Co.

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1):27–52.

Schaffrath, H. (1995). The Essen Folk Song Collection, D. Huron.

Schellenberg, E. G. (1997). Simplifying the Implication-Realization Model of Melodic Expectancy. *Music Perception: An Interdisciplinary Journal*, 14(3):295–318.

Schellenberg, E. G. (2017). Music and Nonmusical Abilities. page 17.

Schellenberg, S. and Moore, R. S. (1985). The Effect of Tonal-Rhythmic Context on Short-Term Memory of Rhythmic and Melodic Sequences. *Bulletin of the Council for Research in Music Education*, (85):207 –217.

Schenker, H. (1935). *Der Freie Satz*, volume III. Universal Edition.

Schumann, R. and Klauser, K. (1860). *Musikalische Haus-Und Lebens-Regeln.* J. Schuberth & Comp.

Smith, M. (1934). Solfège: An essential in musicianship. *Music Supervisors' Journal*, 20(5):16–20.

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. page 93.

Stevens, C. J. (2012). Music Perception and Cognition: A Review of Recent Cross-Cultural Research. *Topics in Cognitive Science*, 4(4):653–667.

Swaminathan, S., Schellenberg, E. G., and Khalil, S. (2017). Revisiting the association between music lessons and intelligence: Training effects or music aptitude? *Intelligence*, 62:119–124.

Talamini, F., Altoè, G., Carretti, B., and Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PLOS ONE*, 12(10):e0186773.

Taylor, J. A. and Pembrook, R. G. (1983). Strategies in memory for short melodies: An extension of Otto Ortmann's 1933 study. *Psychomusicology: A Journal of Research in Music Cognition*, 3(1):16–35.

Unsworth, N., Heitz, R. P., Schrock, J. C., and Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3):498–505.

Wiggins, G. A. (2007). Models of musical similarity. *Musicae Scientiae*, 11(1_suppl):315–338.

Williamson, V. J., Baddeley, A. D., and Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity. *Memory & Cognition*, 38(2):163–175.

Williamson, V. J. and Müllensiefen, D. (2012). Earworms from Three Angles: Situational Antecedents, Personality Predisposition and the Quest for a Musical Formula. *Proceedings from the 12th International Conference on Music Perception and Cognition*, pages 1124–1132.

Wöllner, C. and Halpern, A. R. (2016). Attentional flexibility and memory capacity in conductors and pianists. *Attention, Perception, & Psychophysics*, 78(1):198–208.

Wright, C. R. (2016). Investigating Aural: A Case Study of its Relationship to Degree Success and its Understanding by University Music Students.