

# Modeling Melodic Dictation

*David John Baker*

*2018-09-28*



# Contents

<b>1</b>	<b>Significance of the Study</b>	<b>5</b>
1.1	Claims about need to join the worlds of theory and pedagogy . . . . .	6
1.2	Chapter Overview . . . . .	6
<b>2</b>	<b>Theoretical Background and Rationale</b>	<b>9</b>
2.1	What is melodic dictation? and Why? . . . . .	9
2.2	Individual Factors . . . . .	13
2.3	Musical Factors . . . . .	21
2.4	Modeling and Polymorphism of Ability . . . . .	24
2.5	Conclusions . . . . .	25
<b>3</b>	<b>History of Aural Skills</b>	<b>27</b>
3.1	Thesis: Show that aural skills always has practical end, efficacy of representation of musical pitch . . . . .	28
3.2	Quotes from Schumann . . . . .	28
3.3	Carl Seashore thinking in music . . . . .	28
3.4	Points from Karpinski on pedagogy . . . . .	28
3.5	Points from Royal Paper on pedagogy . . . . .	28
3.6	Solmization System . . . . .	28
3.7	Really this is all question of efficacy of mental representation of musical pitch . . . . .	28
<b>4</b>	<b>Individual Differences</b>	<b>29</b>
4.1	Why care about cognitive abilities . . . . .	30
4.2	Have established that cognitive abilities contribute to musical task (for journal article language repeat) . . . . .	30
4.3	Remind the nature of a musical dictation type task (hear, loop, executive decision) . . . . .	30
4.4	WMC has been misused in music education, theory, pedagogy, aural literature and deserves attention . . . . .	30
4.5	Know WMC plays a role, sense, pertain, execute, should be able to pick up in experiment close to MD . . . . .	30
4.6	Gold-MSI melodic Memory and beat perception test . . . . .	30
4.7	IF we accept these DVs, THEN we should be able to predict them with self reports and measures of WMC and gf . . . . .	30
4.8	Do this with hierarchical LVM ala Elliott paper . . . . .	30
4.9	Overview of Experiment (cross sectional design) . . . . .	30
<b>5</b>	<b>Computation Chapter</b>	<b>31</b>
5.1	Humans like patterns and are very good at picking them up . . . . .	32
5.2	Pre-Musical Corpora . . . . .	32
5.3	Musical Corpora . . . . .	32
5.4	So What? . . . . .	33

<b>6</b>	<b>Hello, Corpus</b>	<b>35</b>
6.1	Brief review of Chapter 4 on corpus (Language to reflect journal submission)	36
6.2	Note problem with using corpus is making corpus	36
6.3	Solem duty to encode and report on corpus	36
6.4	The Corpus	36
6.5	Descriptive Stats of Corpus	36
<b>7</b>	<b>Experiments</b>	<b>37</b>
7.1	Rationale	37
7.2	Experiments	39
7.3	Computational Cognitive Model Model (If time permits) [Whole article in itself]	39
<b>8</b>	<b>Reference Log</b>	<b>45</b>
8.1	To Incorporate	45
8.2	Chapter 3	46

# Chapter 1

## Significance of the Study

All students pursuing 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 hundreds 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 ability 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 competent 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 ubiquity 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 constant 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 in either 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 practitioners of aural skills and people in cognitive psychology CITES. Literature from music theory has established conceptual frameworks regarding aural skills Karpinski (2000) and the relevant cognitive psychology literature has explored factors that might contribute to melodic perception (SCHMUKLER SYNER 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 exactly 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 anecdotal 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 background and rationale for using method from both computational musicology and cognitive psychology in order to answer questions about 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 ability 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 things 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 experiments (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 it 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<sup>1</sup> which is evident given that most aural skills texts that have sections devoted to honing one's listening skills have material on melodic dictation (Karpinski, 2000). Additionally, any school accredited by the National Association of Schools of Music in North America requires learning this skill be part of any accredited curricula as listed under Section 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. 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 even taken a more skeptical stance and have noted 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 that 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<sup>2</sup>, melodic dictation remains at the forefront of many aural

---

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

<sup>2</sup>Of which I can move to later in the thesis

skills classrooms. The act of becoming better at this skill may or may not lead to large increases in far transfer of ability, but used as a pedagogical tool, teaching students to take melodic dictation brings with it concepts that have often been thought of as being at the core of undergraduate music training. While not 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 based on recent work in cognitive psychology and computational musicology. Using this taxonomy, I then review the literature and assert that the next steps forward in understanding how melodic dictation works come from examining the process both experimentally and computationally. It has been nearly two decades since *Aural Skills Acquisition* was first published as the first major step to finally build a bridge between the field of music cognition and music theory pedagogy BUTLER KLONOSKI KARPINSKI ; As with all public works, they need to be updated for the safety of the community?

### 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 his own model of melodic dictation that posits a four step framework.<sup>3</sup>

His four steps include:

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

and occur as a looping process that is depicted in Figure X. 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]

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

---

<sup>3</sup>This four stage process synthesizes 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

Although presumed to be generalizable to however many hearings are played, in the case of the table above, the process happens twice. In his text he presumes this process could happen with a melody that can be extracted in two parts, which according to his prescribed formulas<sup>4</sup> would consist of a melody from twelve to twenty notes long for listeners with few or no chunking abilities. As stated above, reviewing the Karpinski four step model will establish the current state of thinking on melodic dictation and thus create a groundwork from which this current study will build.

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 capacity, may actually play a large 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 X it appears that the stage is not conceptualized as an active process where something like active rehearsal (CITE?) would occur, but rather just consists of 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 evidence 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 (DEUTSCH 1977, (Oura and Hatano, 1988), Handel 1989, Davies 1978, (Dowling, 1990)) Since its publication in 2000, this area of research has expanded with other researchers also demonstrating the effects of musical acculturation via exposure (Eerola, Krumhansl, (Pearce and Wiggins, 2012)).

In describing the short term melodic memory stage, Karpinski also details two processes that he believes to be necessary for this part of the process: extractive listening and chunking. Noting that there is probably some sort of capacity limit to the perception of musical material, citing Miller (Miller, 1956) and corroborating musical evidence (Marple n.d), Karpinski explains how each strategy might be used. Extractive listening is the process in which someone dictating the melody will selectively remember only a small part of the melody in order to ease 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 very complicated concept to actually 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 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 (double check david), but it is worth questioning if it is possible to dissociate 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 translated 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.

<sup>4</sup>In the third chapter of *Aural Skills Acquisition*, Karpinski suggests that the number of playings should be calculated based on the formula FORMULA, where DEFINE WHAT HE SAYS

This last step is ripe for errors and has proved problematic for researchers attempting to study dictation (Taylor and Pembroke, 1983) (Klonoski, 2006). Also worth highlighting is that it is difficult to notate musical material if who is dictating does not have the requisite musical category and knowledge of 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 you would want if you wanted to explore this both experimentally and 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<sup>5</sup>, but there is research from music perception GOLDMAN and other areas of memory psychology STUFF ON CHESS 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 THAT CITATION FROM UP ABOVE. 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 heuristic for communicating the process, this process is generalizable, but intuition would suggest that treating all melodies the same is not going to be 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 X.

[MELODY THAT IS TONAL AND FOLKY]

[MELODY THAT IS BASED ON ATONAL SEQUENCES]

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 will not suffice. This agnosticism for both variability for melodic and individual differences then serves as a stepping off point for this study. In order to have a more thorough understanding of melodic dictation, there needs to be a model that is able to accommodate 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, it will give the community a better sense of the melodic dictation process, which will then enable a more thorough 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 be advantageous for future research to taxonomize the multitude of features thought to contribute to melodic dictation. In doing this, it will 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 X and bifurcates the possible factors thought to affect an individual's ability to take melodic dictation into two areas: **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.

---

<sup>5</sup>And in his Figure 3.1 he does caption it as an *idealized* dictation process

The **individual**, parameters bifurcate 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.<sup>6</sup>

For example, it would be possible to imagine an individual with higher cognitive ability, the opportunity 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 opportunity 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 structural relation of the pitches over time based on performance practice 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 [CITATION?].

This division is not an exhaustive a categorical divide. One could imagine exceptions to these rules where a melody is transformed to the minor key, ornamented, and then played with extensive rubato and experienced as a phenomenologically similar experience.

#### 2.1.1.1 Make a Model of them

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 (refication). These are important to note since from an empirical standpoint, both the task as well as the process of melodic dictation as depicted by Karpinski resemble something that operationalized as both an experiment, as well as a computational model and if understood this way will be subjected to the same 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 (CITE) and are not easily influenced by short term training. 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; ?). Given the strength and generality of these predictors, it is worth investigating the extent that these abilities might contribute when investigating modeling melodic dictation. It is important to understand the degree to which these cognitive factors might influence aural skills abilities in order to ensure that the types of assessments that are given in music school 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 call for a serious upheaval of the current model of aural skills teaching and assessment.

---

<sup>6</sup>Could footnote this interaction and talk about how AP people generally need to have genetic predisposition, fixed pitch instrument in house, and start at early age CITATION FOR THAT?

Recently there has been a surge of interest in this area (DO I CITE THINGS HERE LIKE THAT WORK OF NANCY ROGERS, LEIGH VAN HANDLE, THE UTAH GUY, GARY KARPINSKIS ICMPC, THE FORM AT SMPC) which might be due 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.

Before diving into a discussion regarding differences in cognitive ability, I should note that sometimes ideas regarding differences in cognitive ability been hostilely received (citation against people talking about IQ) and for good reasons. Research in this area can and has been taken advantage to further dangerous ideologies (Bell Curve), 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 fundamental assumption of this study.

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 methodologies that today would be considered risible.

- For example, BROCA thought he could get at intelligence by measuring skulls.
- OR THIS GUY WHO POST HOC/POST MORTUM ASSIGNED GREATNESS RATINGS

While not immediately relevant to current thinking in cognitive psychology, work from both Broca and XXX was continued by the American hereditarian 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 government to identify children struggling in the classroom so that they could be given special attention.

Binet was the initial developer of the idea of an intelligence quotient or IQ<sup>7</sup> 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 Spearman began developing their new theories of intelligence founded on the reification factor analysis. In developing a battery of tests whose performance on one subtest could often reliably predict performance on another— a manifestation of the positive 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 intelligence.

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 there 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 memory; 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 contextual 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 conceptualization manifested into an argument for general crystallized intelligence or  $Gc$ , or DEFINITINO HERE (citation). 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

---

<sup>7</sup>divide mental age by chronological age then multiply by 100

idea conceptualized as general fluid intelligence  $Gf$  (KOVACS AND CONWAY). Even though both of these constructs are powerful predictors on a large scale and do predict things like RICHIE LIST— even when other variables like XYZ are held constant, only conceptualizing cognitive abilities in terms of intelligence does 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.

### 2.2.2 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 relatively small amount of information that one can hold in mind, attend to, or, technically speaking, maintain in a rapidly accessible state at one time. The term working is meant to indicate that mental work requires the use of such information. (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 (?) and later expanded by (?). 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 (?) 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 notable and relevant to this study by 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 that they are stored separately (?) and interpreted in favor of multiple loops (?) , 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 has proposed other models (Cowan, 2005) such as 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 propose 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, two equal-size square bedrooms, and a living room. This is not a bad guess, but it does not rule out the possibility 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) approach, on the other hand, can be compared with hypothesizing that the house includes food preparing quarters, sleeping quarters, bathroom/toilet quarters, and other living

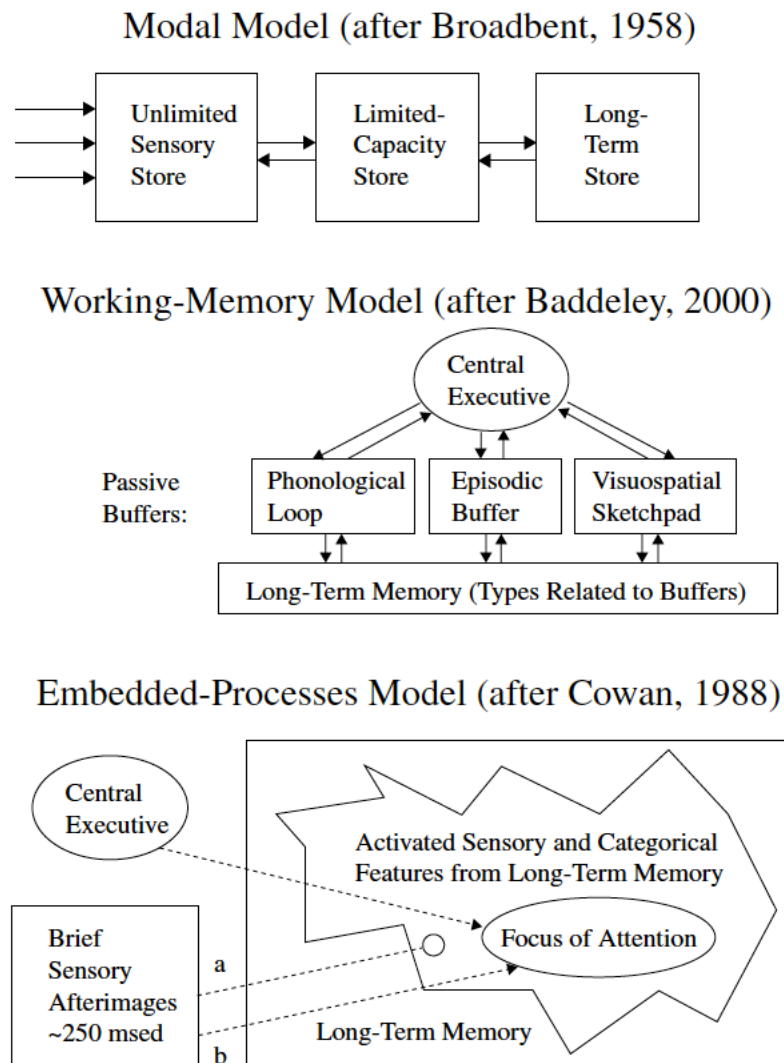


Figure 2.1: Schematics of Models of Working Memory taken from Cowan, 2005



quarters. It is meant to be exhaustive in that nothing was left out, even though it is noncommittal on the details of some of the rooms. p.42 CITATION.

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 who may have not been as critical as they should have been when reading the article. It is worth nothing that Miller has even 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, and has inspired a large amount of research on capacity limits. In the decades since the number 7 has been since reduced to about 4 (?) and research around capacity limits has been investigated using novel methodologies like complex span tasks 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 (?). Given it's predictive strength as well as it's 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 that 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 CITE above. 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 pertinent 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 understood 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 anomalous 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 less possibilities.
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 or the incoming melodic information sounds familiar, 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 inversion 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 is not possible to directly lift work and paradigms from working memory capacity to work in music perception. That said, the enormous amount of theoretical frameworks put forward by the working memory literature when understood in conjunction with theories in music psychology such as implicit statistical learning (Saffran et al., 1999) can provide for new, fruitful theories. Past researchers 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 Baddeley and Hitch modular model of working memory, Berz captures the power of this concept in the last sentence of his article and warns future researchers 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 exhaustively tested since first published in 1994, 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 ability.

### 2.2.2.1 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-analysis by Talamini and colleagues (?) 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-analysis 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 selection bias in engaging with musical activity it also remains a possibility 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 contribute to sight reading tasks based on research where they found measures of working memory capacity, as measured by a matrix span task, to be significantly correlated with many of their measures hypothesized to be related to sight reading ability in pianists at lower difficulty grading (??).

Following up on this work on sight reading, Meinz and Hambrick 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. (?) 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 (?),

that the auditory dictation condition scored surprisingly low and further research might consider further work on dictation abilities. Additionally (?) 2017 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 occurred where the task involved active engagement with novel musical material. This similarity between studies should still be interpreted in light of other research that argued against a need for higher working memory functionality in expert level performances (e.g., HOWE, DAVIDSON, & SLOBODA, 1998).

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 memory abilities.

---

#### 2.2.2.1.0.1 To improve section on next pass

ADD (?)

### 2.2.3 General Intelligence

As discussed above GF has a long history, some good some bad. Thing is that it can be really predictive of many things and people do try to incorporate it into music research. And statistically speaking, terms use IQ, but note that IQ and GF's correlation tend to be statistically identical (Kovacs and Conway, 2016) and many measures of IQ are done with Ravens as seen below.

One of the big problems with this is establishing effects of causality looking at music. Could be that factors outside of music play a role like personal views of ability (Müllensiefen et al., 2015), socio economic status, and personality.

#### 2.2.3.1 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 measures 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 involvement 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 they normally do not reach significance. And also references evidence that when professional musicians are matched with non-musicians in the public, also do not seem to see these associations. Schellenberg, 2015 which Schellenberg takes as more evidence that it is actually just higher functioning kids that take music lessons. 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 methodological problems like how short exposure times were, or researchers who did not hold pre-existing cognitive ability constant (Mehr, Schachner, Katz and Spelke, 2013).

- (Corrigall et al., 2013)
- (?)
- (Swaminathan et al., 2017)

## 2.2.4 Environmental

## 2.2.5 Long term memory and corpus with implicit

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 implicitly held 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 ANDERS ERICKSON 1993 that suggests that performance at more elite levels has to do with 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

Considering points that I make earlier about polymorphic views of musicality and ability.

- Harrison et al for latent variable approach (Harrison et al., 1994)
- Dictation has not been that well researched (Furby, 2016)
- people try lots of things, not much evidence for peer tutoring (Furby, 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 prevalence in school of music curricula. Recently (Fournier et al., 2017) catalogued and categorized 14 different sub categories into four larger main categories while also providing commentary on some of the current state of aural skills. Of the four large categories, they group them into reading mechanisms, sight singing, readings skills acquisition, and learnign support. 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).

Learning to a fluid and competent sight reader presumably bootstraps learning other musical skills, for the same reasons aural skills are taught in schools. Though since it's not directly related and now starting to move away from things looking at dictation directly, time to turn attention to the other half of the taxonomy, the musical factors.

#### CLOSING DISCUSSION ON PUTTING COGNITIVE TOGETHER

- The cognitive stuff if you understand intelligence and WMC as just separate is short sighted, need to incorporate it in some sort of process theory KOVACS AND CONWAY.

## 2.3 Musical Factors

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 be partitioned between both **structural** and **experimental** aspects of a melody. As noted above, there is not a strict delineation between these two categories since one could imagine drastic manipulations in experimental parameters in order to result in a phenomenologically different experience of melody. Questions of transformations of melodies and musical similarity have been addressed in other research CAMPAPERBOULIS and MUSICASCIECNE ARTICLE but are beyond the scope of this study.

### 2.3.1 Structural

The notion that there is something fundamentally inherent and structural to musical material not new to music theory. Heinrich Schenker argued for a hierarchical understanding of tonal music (Schenker, 1935) that asserted a level of hierarchy of notes of the musical surface that has been since expanded upon by theorist throughout the 20th century SALTZER, SCHAKTER, ROTHSTEIN.

Gran quote here about levels of recursion.

The fundamental idea is that there are levels of salience to the music which are perceptual and result from the compositional nature of the notes.

Leonard Meyer in his *Emotion and Meaning in Music* (Meyer, 1956) first tethered the structure that theorists talked about and suggested that in addition to this structure being fundamental to the perception of the piece, were also responsible for some aspects of the emotion and meaning listeners found in the piece. AND ADDITIONALLY SAID THAT WHAT WE THINK OF STYLE IS STATISTICAL PROBABILITY Meyer's work since inspired a line of research investigating the perception of the structural aspects of music could be understood with the work of Eugene Narmour (Narmour, 1990, 1992), Glenn Schellenberg (Schellenberg, 1997), Elizabeth Hellmuth Margulis (Margulis, 2005), David Huron (Huron, 2006) and have inspired computational, machine learning approaches to expectational frameworks in with work by Marcus Pearce (Pearce and Wiggins, 2012). \* WHAT IS THE POINT OF ALL THIS RESEARCH SETNECE

These general models of melodic perception tend to suggest that Meyer was correct in his assertion that computational methodologies could be used to better understand question of melodic perception and structure.

Turning to studies specifically examining melodic dictation with a focus on musical structure, the first study to examine it extensively was Ortmann in 1933 (Ortmann, 1933). Ortmann 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 dictate melodic material. Though Ortmann did not use any statistical methods to model his data, he did determine each of his determinants to contribute to an individual's ability to dictate musical material.

This work was extended by (Taylor and Pembroke, 1983) which additionally incorporated using musical skill as a predictor AND FOUND THAT....

Although the literature is generally sparse 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 methodological decisions such as eliminating 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 parameters thought to contribute like tempo (Hofstetter, 1981) tonality (Dowling, 1978) (Long, 1977) (Pembroke, 1986) (Oura and Hatano, 1988) interval motion (Ortmann, 1933; Pembroke, 1986) length of melody (Long, 1977; Pembroke, 1986) number of presentations (Hofstetter, 1981) [(Pembroke, 1986)] context of presentation (Schellenberg and Moore, 1985) listener experience (Long, 1977; Oura and Hatano, 1988) (Schellenberg and Moore, 1985; Taylor and Pembroke, 1983) familiarity of style (Schellenberg and Moore, 1985)

(Pembroke, 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. DO THEY REPORT TEMPO AND WHAT NOT.

Given the lack of consistent methodologies in administration 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 embedded process model of working memory.

More recently the music education community has also began to do research around melodic dictation using both qualitative and quantitative methodologies. (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 (Pembroke and Riggins, 1990) surveyed various methodologies used by instructors in aural skills settings. Some of these studies consider aural skills as a totality like (Norris, 2003) who provided quantitative evidence to suggest most aural skills pedagogy'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 dictation. They attribute this finding to... encoding differences?

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

Continuing exploring the pedagogical literature, 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 successful strategies that students engaged with when completing melodic dictations. Paney (2016) reported beneficial effects to direct student's attention and guide them through melodic dictation

exercises suggesting that some sort of mental organization of the dictation 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 advantage 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 XXXX Cutberth's Music21 (Cuthbert and Ariza, 2010) which now allows music researchers to systematically digitize symbolic musical material. In addition to creating accessible frameworks for encoding, the computational power available exponentially exceeds that of what was available 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 LIST THEME HERE. An abstracted feature of a melody is a property of the melody... This type of feature abstraction is in contrast to much of the work done in the field of music information retrieval 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**. In the above papers, tended to use more simplistic 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 rigorous computational analyses.

### 2.3.1.2 Static Views of Computational Features/ FANTASTIC

Static features of melodies work by summarizing some aspect of the melody as if it were to be experienced 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 CITATION used in THIS STUDY USING THE ESSEN FOLK SONG ON ARCS 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, ALBRECHT 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 Accessing STatistics (In a Corpus) or FANTASTIC (?). FANTASTIC is software that is capable of summarizing musical material for monophonic melodies. In addition to computing 37 features such as contour variation, tonalness, note density, note length, and measures inspired by computational linguistics (THAT BOOK OF GERAINT), FANTASTIC also calculates m-types (melodic-rhythmic motives) that are based on the frequency distributions of melodic segments found in genres of music. This is inspired by the fact that repetition is key structure of music (Huron, 2006)

Work using the FANTASTIC toolbox has been successful in predicting court case decisions (?), predicting chart successes of songs on the Beatles' *Revolver* (Kopiez and Müllensiefen, 2011), memory for old and new melodies in signal detection experiments (?), memory for earworms (??), memorability of pop music hook (?). 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 assisted platform for studying memory for melodies (?).

### 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 Information 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 expectation for melody IDyOM has been applied to a variety of settings LIST THEM HERE. The domain general application of IDyOM has given credence to Meyer's assertion that the enculturation of musical styles stems from statistical exposure to melodies and be somewhat reflective of the cognitive processes used in musical perception. IDyOM has also been recently extended to look at expectation in multi-part chorales (SAUVE WORK) and expectations of harmony (HARRISON WORK).

(??)

The advantage of using a dynamic approach is that it theoretically reflects real time perception of music with the structural characteristics 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

Remind that the goal here is to create a model. Problem is a lot of the things discussed in the literature review are abstract concepts. Tonalness, musicianship, intelligence, working memory, even aural skills and can not be measured directly

And on top of that we have the idea about levels of Explanation - SEE THAT WITH WIGGINS

Important here to note that unlike Harrison 1994, not really appropriate to conceptualize this as latent ability. Very easy in statistical terms to talk about latent variables and be able to talk about lots of variance predicted. But problem with any sort of model like that or model of musical sophistication is that it has same fate as g.

QUOTES HERE FROM MS paper on latent variables

Musicianship as a concept of something to measure when talking about this one little task with relative lack of concrete evidence for its transfer, need to return back to a process model of melodic dictation.

### 2.4.1 Polymorphic, component process makes you think about things in models

Going to say here that LVs are bad and that need to adopt a more polymorphic view of musicianship as advocated by

- (Levitin, 2012; Peretz and Coltheart, 2003)



## 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 comprehensive 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 systemize 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

- Dowling 1991 (Dowling, 1991)





## Chapter 3

# History of Aural Skills

### 3.1 Thesis: Show that aural skills always has practical end, efficacy of representation of musical pitch

3.1.1 for i in star aural people do

3.1.2 Who

3.1.3 Where

3.1.4 When

3.1.5 What

3.1.6 How (approach and goals)

3.1.7 Why

3.1.8 Guido d'Arezzo

3.1.9 Walerant (via Calvisius)

3.1.10 Banchieri

3.1.11 Cerratto

3.1.12 Penna

3.1.13 Zarlino

### 3.2 Quotes from Schumann

### 3.3 Carl Seashore thinking in music

### 3.4 Points from Karpinski on pedagogy

### 3.5 Points from Royal Paper on pedagogy



## Chapter 4

# Individual Differences

### 4.1 Why care about cognitive abilities

#### 4.1.1 General intelligence and WMC

#### 4.1.2 Defining of terms

### 4.2 Have established that cognitive abilities contribute to musical task (for journal article language repeat)

#### 4.2.1 General Fluid Intelligence, WMC, Training as uni of polymorphic

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

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

##### 4.3.1.1 Berz 1994 noticed it first

##### 4.3.1.2 Williamson Baddely Hitch suggest maybe musical loop

##### 4.3.1.3 Even Cowan labs wonder how different (Li Cowan Saults )

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

#### 4.4.1 Problems with chunking

##### 4.4.1.1 Mistake with Miller 1956, he did not mean 7 items

##### 4.4.1.2 Broadbent 1956 more of why its more like 3-4

#### 4.4.2 Problems with using capacity limit literature

##### 4.4.2.1 See Cowan 2005 page 80

##### 4.4.2.2 Musical order is always serial effects

## Chapter 5

# Computation Chapter

OTHER PEOPLE WHO HAVE DONE THIS

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 et al 2007 and 2007
- McCay 2005
- Huron 2006
- Frieler 2008
- JAZZOMAT PROJECT OUTPUT

## 5.1 Humans like patterns and are very good at picking them up

### 5.1.1 We learn things implicitly

### 5.1.2 We can represent that implicit knowledge with a corpus

## 5.2 Pre-Musical Corpora

### 5.2.1 Information Theory

### 5.2.2 Computational Linguistics as front runner

## 5.3 Musical Corpora

### 5.3.1 History of Musical Corpora

#### 5.3.1.1 Fun old computational music papers

#### 5.3.1.2 Corpora that are often used

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

### 5.3.2 FANTASTIC

#### 5.3.2.1 static

#### 5.3.2.2 ML approach gets it right

#### 5.3.2.3 simple to understand

#### 5.3.2.4 Can abstract features be perceived?

##### 5.3.2.4.0.1 Note density

##### 5.3.2.4.0.2 Contour variation

##### 5.3.2.4.0.3 Tonalness

##### 5.3.2.4.0.4 weird computational measures

### 5.3.3 IDyOM as representation of musical materials

#### 5.3.3.1 n-gram models

#### 5.3.3.2 mirrors human behavior

##### 5.3.3.2.0.1 melody



5.3.3.2.0.2 harmony

## 5.4 So What?

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

5.4.0.2 Can operationalize item level items contextually with a corpus

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

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

5.4.0.5 We need new materials





## Chapter 6

# Hello, Corpus

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

#### 6.1.1 Corpus outside of music

#### 6.1.2 Corpus in Music

#### 6.1.3 The point is that it implicitly represents humand knowledge

#### 6.1.4 IDyOM 1

#### 6.1.5 IDyOM 2

#### 6.1.6 IDyOM 3

#### 6.1.7 Huron suggestions that starts of melodies relate to mental rotaiton

#### 6.1.8 Other Huron claims

### 6.2 Note problem with using corpus is making corpus

#### 6.2.1 Many are used on Essen

#### 6.2.2 Brinkman says Essen Sucks

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

### 6.3 Solem duty to encode and report on corpus

#### 6.3.1 Justin London Article on what makes it into a corpusu

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

### 6.4 The Corpus

#### 6.4.1 History of Sight Singign books

# Chapter 7

## Experiments

### 7.1 Rationale

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

7.1.2 Clearly many factors contribute to this whole thing and need to be taken into a model

7.1.3 Dictation is basically a within subjects design Experiment

7.1.3.1 Get very ecological and dirty and run it

- Paney 2016 had 30 second timing

### 7.1.4 Factors

7.1.4.1 Cognitive

7.1.4.1.1 WMC

7.1.4.1.2 GF

7.1.4.2 Training

7.1.4.2.1 Goldsmiths MSI

7.1.4.3 Musical

7.1.4.3.1 FANTASTIC

7.1.4.3.2 IDyOM



7.1.4.4 Investigate melodies with this context and set scoring

7.1.4.5 Mirror design to see if effects of melody are there

## 7.2 Experiments

### 7.2.1 Experiment I

7.2.1.1 Participants

7.2.1.2 Procedure

7.2.1.3 Materials

7.2.1.4 Scoring

7.2.1.5 Results

7.2.1.6 Modeling

7.2.1.7 Discussion

### 7.2.2 Experiment II

7.2.2.1 Participants (New)

7.2.2.2 Procedure (Same)

7.2.2.3 Materials (Swapped but controlled)

7.2.2.4 Scoring (Same)

7.2.2.5 Results

7.2.2.6 Modeling (same)

### 7.2.3 General Discussion

7.2.3.1 What happened

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

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

7.2.3.4 Also how we score it is going to mess with the DVs

7.2.4 Really what is needed is Computational Model

## 7.3 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 purview 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 dictation 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 every 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 GGLOT2. 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 be 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 succeeds 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 the threshold 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 threshold, 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 guesses 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 Karpinski rails against 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 recency 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 phenomenological experience of taking melodic dictation.

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?

### 7.3.1 Why?

#### 7.3.1.0.1 Better than verbal models

#### 7.3.1.0.2 Sometimes even mathematically infesable proposed theory

#### 7.3.1.0.3 Beyond Karpinski in that it doesn't just schematize, says exactly when each thing is happening when

#### 7.3.1.0.4 Lends itself to better discussions that don't just rely on personal anecdotes

#### 7.3.1.0.5 Can tweak the parameters

#### 7.3.1.0.6 Can collect different types of data (corpus or experimental) and use the model

#### 7.3.1.0.7 This model suggests that atomism approach is actual just subprocess of larger pattern

### 7.3.1.1 Theoretical Justification

7.3.1.1.1 Marries literature on LTM and prior knowledge, information theory, WMC, computation, representation

7.3.1.1.2 Also can be implemented in computer

7.3.1.1.3 representation of rhythm too?

7.3.1.1.4 inspired by people like margulis 2005, albrecht and shanahan key finding, want something to contribute

7.3.1.1.5 Really Made me think

7.3.1.2 The Model (note many parameters can be changed in R package)

7.3.1.3 Prior

7.3.1.3.0.1 Corpus of music represented in form of n-grams

7.3.1.3.0.2 IDyOM extracts all possible n-gram permutations as learned corpus

7.3.1.3.1 Music notation fed into processing window where incoming n-gram is matched based on WMC window OR IT maximum

7.3.1.3.1.1 Information builds until approaches critical threshold

7.3.1.3.1.2 Upon maximum, model puts n-gram into focus of attention (Cowan 1988) and note why this is better than Baddely Hitch

7.3.1.3.1.3 Recursive transcribe function looks for LTM matches

Option 1: Pattern Matched and Pattern Transcribed, success?

Option 2: Pattern not matched in full, truncated and use match option again (should be higher probability of match with corpus)

Option 3: Pattern not matched downsize again until at interval level and relying on 2-gram (atomism)

On success of option, reopen gate at nearest long implicit n-gram LTM Match (start or end problem)

7.3.1.3.1.4 Put time constraints on search features

7.3.1.3.1.5 Transcribe process resets with trace image of melody after each dictation

7.3.1.3.1.6 Transcribe process ends when all notes accounted for

### 7.3. COMPUTATIONAL COGNITIVE MODEL MODEL (IF TIME PERMITS) [WHOLE ARTICLE IN ITSELF]<sup>43</sup>

#### 7.3.1.4 Model Output

7.3.1.4.1 Based on learning, times needed to hear it

7.3.1.4.2 Completion percentage

7.3.1.4.3 Rank order of easier to transcribe parts based on learning

#### 7.3.1.5 Model Compared to Data

7.3.1.5.1 With Experimental Data

### 7.3.2 Future Suggestions for Aural Skills Pedagogues and Research

7.3.2.1 Use model as teaching stepping off point

7.3.2.2 Should move towards LTM pattern matching

7.3.2.3 Reason that people learn how to sight sing is to INCREASE the learning of the implicit corpus

7.3.2.4 Circular process here

7.3.2.5 Is this what it means to then think IN music

7.3.2.6 Really it's to just know the patterns maybe like model where Justin London suggests we get to know patterns and expect them

7.3.2.7 Would also make sense in terms of Leonard Meyer 1956

7.3.2.8 Use WMC in music theory, cognition, education studies



## 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 structural 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 recommend things for sight singing
- (Pembrook, 1986) – Effects of melodies, also how people do it. Interesting that they too effect of melodies, but talk about things in terms of notes and not in terms of information content. Thought of 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 ability 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
- (Pembroke 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 dictation 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 Baddeley 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 citation 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.
- 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.
- 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.
- 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).
- 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. (2013). Music Training, Cognition, and Personality. *Frontiers in Psychology*, 4.
- Cowan, N. (2005). *Working Memory Capacity*. Working memory capacity. Psychology Press, New York, NY, US.

- 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.
- 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.
- Gould, S. J. (1996). *The Mismeasure of Man*. WW Norton & Company.
- 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-Baased 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. (2006). *Sweet Anticipation*. MIT Press.
- 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 Müllensiefen, 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 Popularität. Beiträge 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.
- 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.
- 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, 1. gardner edition.
- 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.
- 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. and Wiggins, G. A. (2012). 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 Successful Dictation Strategies. *Journal of Music Theory Pedagogy*, 4(1):63–72.
- 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.
- 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.
- 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.
- Taylor, J. A. and Pembroke, 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.