# Modeling Melodic Dictation

*David John Baker*

*2019-03-15*

# Contents


**1 Significance of the Study** **5**
- 1.1 Rationale . . . 5
- 1.2 Chapter Overview . . . 6
- 1.3 Dissertation Output . . . 7

**2 Theoretical Background and Rationale** **9**
- 2.1 Melodic Dictation . . . 9
- 2.2 Individual Factors . . . 14
- 2.3 Musical Parameters . . . 22
- 2.4 Modeling and Polymorphism of Ability . . . 25
- 2.5 Conclusions . . . 26

**3 Individual Differences** **27**
- 3.1 Rationale . . . 27
- 3.2 Individual Differences . . . 28
- 3.3 Overview of Experiment . . . 32
- 3.4 Discussion . . . 37

**4 Computation Chapter** **45**
- 4.1 Rationale . . . 45
- 4.2 Agreeing on Complexity . . . 46
- 4.3 Modeling Complexity . . . 56
- 4.4 Frequency Facilitation Hypothesis . . . 62
- 4.5 Conclusions . . . 68

**5 Hello, Corpus** **69**
- 5.1 Rationale . . . 69
- 5.2 History . . . 70
- 5.3 Berkowitz Corpus . . . 71
- 5.4 Comparison of Corpora . . . 71

**6 Experiments** **73**
- 6.1 Introduction . . . 73
- 6.2 Methods . . . 76
- 6.3 Results . . . 77
- 6.4 Discussion . . . 78

**7 Computational Model** **79**
- 7.1 Levels of Abstraction . . . 79
- 7.2 Model Overview . . . 80
- 7.3 Verbal Model . . . 80
- 7.4 Formal Model . . . 85
- 7.5 Conclusions . . . 87

**8 Reference Log** **89**
8.1 To Incorporate . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89
8.2 Chapter 3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 90

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.1.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.8
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.3.1     v forcats 0.3.0

## -- Conflicts ------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

# Chapter 1

# Significance of the Study

## 1.1 Rationale

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 (**?**, §VIII.6.B.2.A). Melodic dictation is a cognitively demanding process that requires students to hear a melody, then without any access to an external reference, transcribe the melody on paper within a limited time frame. As of 2019 there are 644 Schools of Music belonging to National Association of Schools of Music (NASM)[1], meaning that thousands of students every year will be expected to learn this challenging task as part of their aural skills education. The implicit logic is that as one improves in their ability to take melodic dictation, this practice of critical and active listening develops one's ability to "think in music" (**??**) and thus become a more competent musician.

Despite its ubiquity in curricula within School of Music settings, research that explains how people learn melodies is limited at best. 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 these subjects do not overlap in formal training, findings related to this question are published in separate journals, and thus the overlapping research is scarce. This problem is not new and there have been repeated attempts to bridge the gap between practitioners of aural skills and researchers in cognitive psychology (**??????**). Literature from music theory has established conceptual frameworks regarding aural skills (**?**)[2], cognitive psychology literature has explored factors that might contribute to melodic perception (**????**), and applied literature from the field of music education (**???**) has investigated how people learn melodies in more ecological settings.

However, despite these isolated areas of research, we as music researchers, do not have an a concrete understanding of exactly what contributes to the process of how individuals learn melodies (**?**). 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. This chasm in the literature also raises a disconcerting question in music pedagogy:

> If we as pedagogues do not have a in-depth understanding of how people learn melodies, how can we fairly assess what students can be expected to accomplish in the classroom and then fairly grade them on their attempts?

While no single dissertation can solve any problem completely, this dissertation aims to fill the gap in the literature between aural skills practitioners and music psychologists in order to reach conclusions that can be applied systematically in pedagogical contexts. I do this by synthesizing literature from both music theory and music cognition in order to demonstrate how tools from both cognitive psychology as well as computational musicology can be used to help inform pedagogical practices.

[1] For a list of Schools, please look here
[2] More aural skills textbooks here

## 1.2 Chapter Overview

In the first chapter, I begin by building off of the work of Gary Karpinski (**??**) in order to introduce the process of melodic dictation and discuss factors that might play a role in a student's ability to take melodic dictation. The chapter introduces both a theoretical background and rationale for using methods from both computational musicology and cognitive psychology in order to answer questions about how individuals learn melodies. In order to organize the disparate literature, I put forward a taxonomy of factors that are assumed to contribute to an individual's ability to take melodic dictation and discuss each in turn. This chapter outlines the factors hypothesized to contribute to an individual's ability to learn melodies, incorporating both individual and musical parameters. I conclude the chapter with a discussion of some philosophical and theoretical problems when attempting to measure issues concerning melodic dictation and argue for the advantages of answering this problem using a polymorphic view of musicianship (**???**).

The second chapter of the dissertation investigates individual factors that are theorized to contribute to melodic dictation. I argue that since the first two steps of Karpinski's model of melodic dictation do not require any musical training, teasing apart the individual factors that contribute to melodic dictation can be done using a memory for melodies paradigm. I interpret the results of an experiment to highlight the importance of working memory processes in melodic dictation. The chapter corroborates claims by **?** on the importance of understanding differences in working memory capacity and establishes rationale for including it as a variable of interest in future research on melodic dictation.

The third chapter of the dissertation discusses how aural skills pedagogy could benefit from using methodologies from computational musicology in order to inform their teaching practice. The chapter begins by establishing the degree to which aural skills pedagogues agree on the difficulty of melodies for melodic dictation using a survey representing 40 aural skills pedagogues. I then show how different sets of tools from computational musicology can approximate the intuitions of aural skills pedagogues using the survey data as a ground truth. The chapter concludes by putting forward a novel theory of musical memory– The Distributional Facilitation Hypothesis– which combines theoretical work from cognitive psychology and computational musicology. I show how this hypothesis can be applied in pedagogical settings to create a more linear path to success in the aural skills classroom for students.

In my fourth chapter, I introduce a novel corpus of 783 digitized melodies encoded in the `**kern` format (**?**). 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 discuss how the corpus could be used to generate hypotheses for future experiments. This dataset serves as a valuable resource for future researchers in music, psychology, and the digital humanities.

In the fifth chapter, I synthesize the previous research in a melodic dictation experiment. Stimuli for the experiments are selected based on the symbolic features of the melodies discussed in earlier chapters and are manipulated as the independent variables. I then model responses from the experiments using both individual factors and musical features using mixed-effects modeling in order to predict how well an individual performs in behavioral tasks. In discussing the results, I also note important caveats in scoring melodic dictation and highlight how changes in scoring can lead to changes in the final modeling. Results from this chapter will be discussed with reference to how findings are applicable to pedagogues in aural skills settings.

Finally, in my sixth chapter, I introduce a computational, cognitive model of melodic dictation with the goal of helping explain how students improve at melodic dictation. The model is based in research from both cognitive psychology (**?**) and computational musicology (**?**) and incorporates relevant theoretical aspects such as working memory (**??**) and the structure of the melody itself. In this chapter I demonstrate how modeling the cognitive decision process during melodic dictation helps provide a precise framework for pedagogues to understand student's inner cognition during melodic dictation and can help inform teaching practice.

## 1.3 Dissertation Output

### 1.3.1 Reserach Papers

- SEM Paper in ICMPC
- Aural Skills Paper ICMPC
- Aural Skills Textbook
- Goldsmiths Replication

### 1.3.2 Research Presentations

- SEM Paper at ICMPC
- Aural Skills Poster
- Milton Keynes Talk
- Public Musicology
- Again and Again?

# Chapter 2

# Theoretical Background and Rationale

## 2.1 Melodic Dictation

Melodic dictation is the process in which an individual hears a melody, retains it in memory, and then uses their knowledge of Western musical notation to recreate the mental image of the melody on paper in a limited time frame. For many, becoming proficient at this task is at the core of developing one's aural skills (**?**). For over a century, music pedagogues have valued melodic dictation[1] which is evident from the fact that most aural skills texts with content devoted to honing one's listening skills have sections on melodic dictation (**?**).

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

Some researchers have taken a more skeptical stance and asserted that the rationale for why we teach melodic dictation deserves more critique. For example, Klonoski in writing about aural skills education aptly questions "What specific deficiency is revealed with an incorrect response in melodic dictation settings?" (**?**). Earlier researchers like Potter, in their own publications, have noted how they have been baffled that many musicians do not actually keep up with their melodic dictation abilities after their formal educaiton ends (**?**), 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 (**?**), seems somewhat exclusionary to musical cultures that do not depend on any sort of written notation.

Though despite this skepticism towards the topic, melodic dictation remains at the forefront of many aural skills classrooms. The act of becoming better at this skill may or may not lead to large in increases in far transfer of ability, but used as a pedagogical tool, teaching students to take melodic dictation brings with it concepts that have been deemed relevant to the core of undergraduate music training. While there has not been extensive research on melodic dictation research in recent years– in fact **?** notes that since 2000, only four studies were published that directly examined melodic dictation– this skill set sits on the border

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

between literature on learning, melodic perception, memory, and music theory pedagogy. Understanding and modeling exactly how melodic dictation works remains as a untapped watershed of knowledge for the field of music theory, music education, and music perception and is deserving of much more attention.

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

### 2.1.1 Describing Melodic Dictation

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

The four steps of Karpinski's model include

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

and occur as a looping process which I have reproduced depicted in Figure 2.1. The model is discussed extensively in both the original article (**?**) and throughout the third chapter in his book (**?**).

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

The **short-term melodic memory** stage in this process references the point in a melodic dictation where musical material is held in active memory. From Figure 2.1 it appears that the stage is not conceptualized as an active process where something like active rehearsal would occur, but rather just consists soley of passive mental representation. Though Karpinski does not posit any sort of active process in the short term melodic memory stage, he does suggest there are two separate memory encoding mechanisms, one for contour, and one for pitch. He arrives at these two mechanisms by using both empirical qualitative interview evidece as well as noting literature from music perception that supports this claim for contour (**??**) and literature suggesting that memory for melodic material is dependant on enculturation (**???**). Since its publication in 2000, this area of research has expanded with other reearchers also demonstrating the effects of musical acculturation via exposure (**???**).

In describing the short term melodic memory stage, Karpinksi also details two processes that he believes to be nesscary for this part of melodic dictation: extractive listening and chunking. Noting that there is probably some sort of capacity limit to the perception of musical material, citing Miller (**?**), Karpinski explains how each strategy might be used. Extractive listening is the process in which someone dictating the melody will

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

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

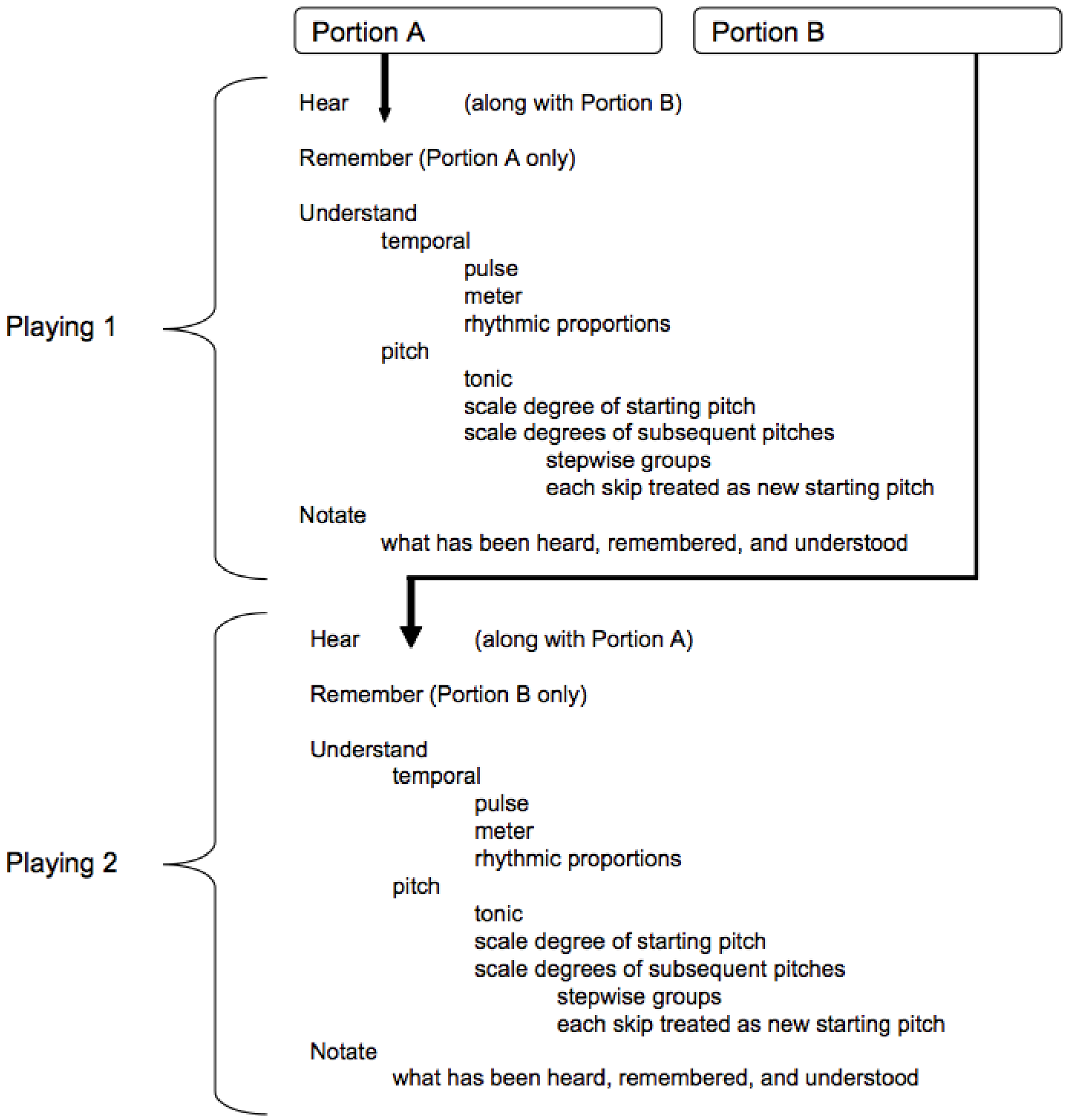


Figure 2.1: Karpinski Idealized Flowchart of Melodic Dictation

selectivly remember only a small part of the melody in order to lessen the load on memory. Chunking is the process in which smaller musical elements can be fused together in order to expand how much information can be actively held in memory and manipulated. The concept of chunking is very helpful as a pedagogical tool, but as detailed below, is a complicated concept to pin down how it works.

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

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

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

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

Second, the model presumes the same sequence of events for every melody. As a general heurstic for communicating the process, this process is generalizable, but intuition would suggest that treating all melodies the same is not going to lead to having a robust model. For example, on page 103, Karpinski suggest that two listenings should be adequate for a listener with few to no chunking skills to listen to be able to dictate a melody of twelve to twenty notes. This process might generalize to many tonal melodies, but presumably different strategies in recognition would be involved in dictating the two melodies of equal length shown in Figure 2.2 and 2.3.

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

[4]And in his Figure 3.1 he does caption it as an *idealized* dictation process

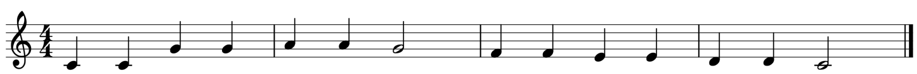

Figure 2.2: Tonal Melody

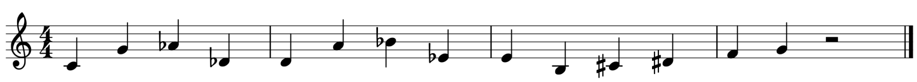

Figure 2.3: Atonal Melody

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

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

The taxonomy that I propose appears in Figure 2.4 and bifucates the possible factors thought to affect an individual's ability to take melodic dictation into two categories: **individual** parameters and **musical** parameters. Each of these two categories can then be split again into **cognitive** and **environmental** parameters as well as **structural** and **experimental factors** respectively. Below I expand on what these categories entail, then explore each in depth.

The **individual** parameters split broadly into **cognitive factors**, or factors of people that are relatively consistent with people over time and could be understood as largely being governed by *nature*. The other category of this division consist of factors that change with training and exposure and could be understand as largely being governed by *nurture*. This second set of parameters are the **environmental** factors. These categories are not deterministic, nor exclusive, and almost inevitably interact with one another.[5]

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

Complementing the individual differences, there would also be differences at the **musical** level which in turn divides into two categories. On one hand exists the **structural aspects** of the melody itself. These are aspects of the melody that would remain invariant when written down on a score. Parameters in this

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

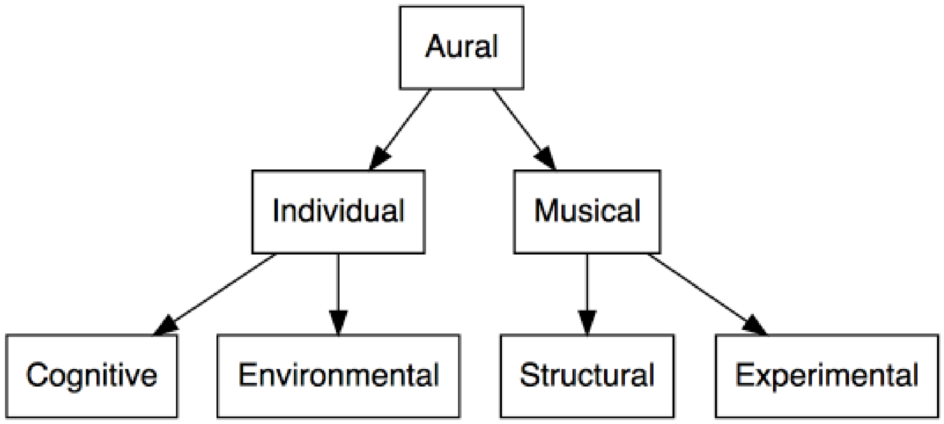


Figure 2.4: Taxonomy of Factors Contributing to Aural Skills

category would include features generated by the interval structure of the pitches over time that allow the melody to be categorically distinct from other melodies. These structural features are then complimented by the **experimental features** which are emergent properties of the structrual relation of the pitches over time based on performance pratice choices. Examples of these parameters would include, key, tempo, note density, timbral qualities, and the amount of times a melody is played during a melodic dictation or emergent properties like a melody's tonalness as computed through various metrics. This division is not an exhaustive, categorical divide. One could imagine exceptions to these rules where a melody is tranformed to the minor key, ornamented, and then played with extenive rubato and experienced as a phenomenologically similiar experience.

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

## 2.2 Individual Factors

### 2.2.1 Cognitive

- (**?**) – musical training, rather than WM predicts faking a second langauge in music

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

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

Before diving into a discussion regarding differences in cognitive ability, I should note that sometimes ideas regarding differences in cognitive ability been hostily received (citation against people talking about IQ) and for good reasons. Research in this area can and has been taken advantage to further specious ideologies, but often arguments that assert meaningful differences in cognitive abilities between groups are founded on statistical misunderstandings and have been debunked in other literature (**?**). Considering that, it then becomes very difficult to maintain a scientific commitment to the theory of evolution (**?**) 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 desereves to be examined since the ability to validly and reliably measure an individual's cognitive ability is a fundmental assumption of this study.

### 2.2.2 Measuring Intelligence

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

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

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

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

Though seemingly unrelated to the current state of thinking about cognitive abilities, Binet's and Spearman's ideologies about what intelligence is and how to measure it still represent two of the larger schools on

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

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

cognitive ability. On one hand their idea that cognitive abilities are based upon a steady growth of incoming information that someone is able to manipulate once they retrieve from long term memor; on the other hand there is a school of thought that there is some sort of measureable construct, *g* that aids in the process of solving problems that do not depend on any sort of contextuxal information. Conceptualizing cognitive ability as these two different constructs inevitably leads to different types of measurements and subsequently what these constructs are then able to predict in terms of human behavior.

Without detailing entire histories of both lines of thought, Binet's conceptulization manifested into an argument for general crystallized intelligence or *Gc*, or the ability to solve problems based on previously acquired skills. Spearman and Burt's ideas about *g* school reflect a belief that individuals have some sort of latent cognitive ability to draw on to perform mental tasks. The cognitive psychology literature has noted that *g* often shares a statistically equivalent relationship to idea conceptualized as general fluid intelligence *Gf*, or the ability to solve problems in novel situations Cattell, 1971; Horn, 1994). This distinction between *Gf* and *Gc* is different than that of *g*, but again it should be noted that *Gf* and *g* share a statistically identical relationship Matzke, Dolan, and Molenaar (2010). These conceptions of intelligenec and cognitive ability aslo differ from more current theories that sythesize these previous areas of research (**?**).

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

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

### 2.2.3 Working Memory Capacity

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

> the 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 mean to indicate that mental work requires the use of such information. (p.1) (**?**)

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 (**?**), one of the first modal models of memory was proposed by (**?**) and later expanded by (**?**). As seen in Figure 2.5, 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 (**?**) 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 relevant to this study is by (**?**), who postulated adding a musical rehearsal loop to the already established phonological loop and visual spatial sketchpad. While Berz is most likely correct in asserting that the nature of storing and processing musical information is different to that of words or pictures and there has been experimental evidence to suggest this (**?**) that has been 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 proposed a separate model (**??**) dubbed the Embedded Process Model which do not claim the existence

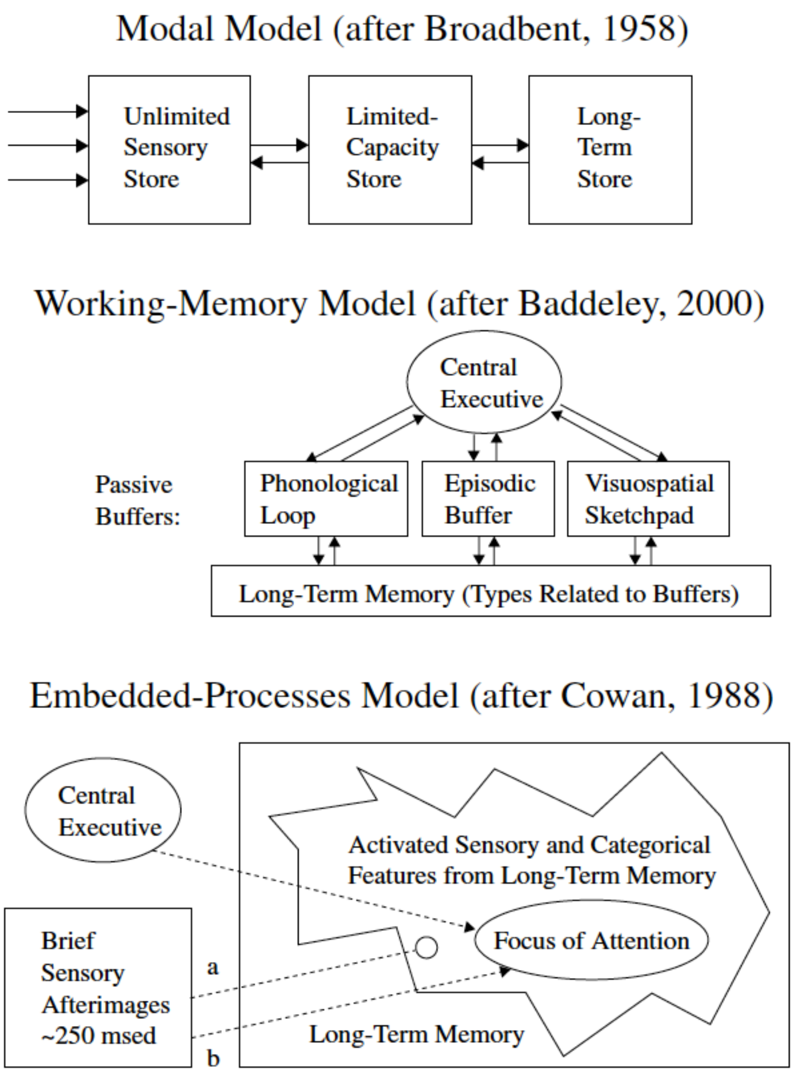


Figure 2.5: Schematics of Models of Working Memory taken from Cowan, 2005

of any domain specific module (e.g. positing a phonological loop, visual spatial sketchpad) but is rather based on an exhaustive model that did away with the problem of asserting specific buffers for new types of information.

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

> The aim was to see if the description of the processing structure could be exhaustive, even if not complete, in detail. By analogy, consider two descriptions of a house that has not been explored completely. Perhaps it has only been examined from the outside. Baddeley's (1986) approach to modeling can be compared with hypothesizing that there is a kitchen, a bathroom, two equal-size square bedrooms, and a living room. This is not a bad guess, but it does not rule out the possibllity that there actually are extra bedrooms or bathroom, that the bedroom space is apportioned into two rooms very different in size, or that other rooms exist in the house. Cowan's (1988) apporach, on the other hand, can be compared with hypothesizing that the house includes food preparating quarters, sleeping quarters, batroom/toilet quarters, and other living quarters. It is meant to be exhaustive in that nothing was left out, even though it is noncommital on the details of some of the rooms. p.42 Cowan, 2005.

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

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

#### 2.2.3.1 Parallels Between Working Memory and Melodic Dictation

Given its predictive strength as well as its direct similarity to tasks of melodic dictation, a in depth look at the literature is warranted.
Clearly an individual's ability to take in sensory information, maintain it in memory, actively carry out other tasks (like notating said melody) are almost identical to tasks of working memory capacity. Before venturing onward from this striking parallel, tasks investigating working memory capacity differ from melodic dictation tasks in a few key ways. The first is that musical information is always sequential: a melodic dictation task would never require the student to recall the pitches back in scrambled orders. Serial order recall is an important characteristic in the scoring and analyzing of working memory tasks (**?**), but musical tones do not appear in random order and are normally in discernable chunks as discussed by Karpinski(**?**). 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 (**?**), and any pertitant music psychology literature.

1. *Chunks may have a hierarchical organization.* Tonal music has historically been understood to be hierarchical (**???**) with the study for memory for tones being confounded by some pitches being

undestood by their relation to structurally more important tones.

2. *The working memory load may be reduced as working memory shifts between levels in hierarchy.* If an individual understands a chunk to be something such as a major triad, the load on working memory would be less since it that information could be understood as a singular chunk.
3. *Chunks may be the endpoints of a continuum of associations.* Pairing a group of tones together that might be functionally anomolous like...
4. *Chunks may include asymmetrical information.* More tonal possibilities are possible from a stable note like tonic or dominant, whereas in a tonal context, a raised scale degree $\#\hat{4}$ when understood in a functional context would be taken as having stricter transitional probabilities ($\#\hat{4} \rightarrow \hat{5}$).
5. *There may be a complex network of associations.* If a set of pitches sounds like a similar set of pitches from long term memory , the information coming in can not be understood as being separate units of working memory.
6. *Chunks may increase in size rapidly over time.* Three tones that are seemingly unrelated when incoming like E4, G5, C5 might enter sensory perception as three different tones, but then be fused together when they are understood as one chunk– a first inverstion major triad.
7. *Information in working memory may benefit from rapid storage in long term memory.* Given the amount of patterns that an individual learns and can understand, as soon as something is fused, it could be encoded in long term memory, especially if there is a salient feature in the incoming melodic information such as the immediate recognition of a mode or cadence.

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

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

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

#### 2.2.3.2 Working Memory Capacity and Music

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

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

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 (**?**) 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 (**?**) found that working memory capacity, as measured by a backwards digit span and operation span, to be successful predictors in a tapping task requiring sensory motor prediction abilities. As mentioned above, each of these tasks where working memory was a significant predictor of performance occured where the task involved active enagement with novel musical material.

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

### 2.2.4 General Intelligence

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

#### 2.2.4.1 Papers that suggest GF plays a role

- Redo this structure

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

- (**?**)
- (**?**)

### 2.2.5 Environmental

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

### 2.2.6 Musical Training

- Papers that suggest practicing makes you better?
- It almost seems redundant to review literature in support of music practice leading to better results.
- List of those papers here

### 2.2.7 Aural Training

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

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

### 2.2.8 Sight Singing

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

FIX HERE

The authors note a line of research that has documented that university students are often unprepared to sight-read single lines of music (Asmus, 2004; Davidson, Scripp & Welsh, 1988; Fournier, 2015; Thompson, 2004; Vujović & Bogunović, 2012) even though it is, like dictation, thought of as a means for deeper musical understanding. (DeBellis, 2005; Karpinski, 2000; Ottman, 1956; Rogers, 2004; Scripp, 1995; Scripp & Davidson, 1994) The authors of Fournier et. al also note that sight-reading has been an active area of research due to the often reported relationship that performance on sight reading often predicts several studies have shown links between academic success in sight-singing and predictors such as entrance tests (Harrison, 1987, 1990, 1991; Ottman, 1956; Rodeheaver, 1972; Schleuter, 1983), academic ability (Chadwick, 1933; Harrison, 1990, 1991; Harrison, Asmus, & Serpe, 1994; Rodeheaver,1972), and musical experience (Brown, 2001; Dean, 1937; Furby, 2008; Harrison, 1990, 1991; Harrison et al., 1994; Thostenson, 1967).

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

## 2.3 Musical Parameters

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

### 2.3.1 Structural

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

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

- WHAT IS THE POINT OF ALL THIS RESERACH SETNECE

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

[8]This whole section sucks

Turning to studies examining melodic dictation with a focus on musical structure, the first study to examine it extensively was Ortmann in 1933 (**?**). Ortamnn used a series of twenty five-note melodies in order to examine the effects of repetition, pitch direction, conjunct-disjunct motion (contour), interval size, order, and chord structure, all of which he deems to be the *determinants* of an individual's ability to dictate melodic material. Though Ortmann did not use any statistical methods to model his data, he did assert that each of his determinants contributed to an individual's ability to dictate musical material. This work was extended by (**?**) which additionally incoporated using musical skill as a predictor and additionally found evidence that these factors contributed to individual dictation abilities in a sample of 122 undergraduate students.

Although the literature is generally sparce compared to other areas of music cognition, literature exploring the effects of structural characteristics on memory does exist. Long found that length, tonal structure, contour, and individual traits all contribute to performance on melodic dictation examinations and found that structure and tonalness to have significant, albeit small predictive powers in modeling (**?**). One problem with studies such as (**?**) is that they sometimes would make conspicuous methodloogical decisions such as elimnating individuals who were bad singers for the example. Not only does this reduce the spectrum of ability levels (assuming that singing ability correlates with dictation ability, a finding since which has been established (**?**)), but is additinally 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 (**?**).

Other researchers have also put forward other paramters thought to contribute like tempo (**?**) tonality (**?**) (**?**) (**?**) (**?**) interval motion (**??**) length of melody (**??**) number of presentations (**?**) [(**?**)] context of presentation (**?**) listener experience (**??**) (**??**) familarity of style (**?**)

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

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

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

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

Continuing exploring the pedagogical liteature, Naton Buonviri and colleagues have also made melodic dictation a central focus of some recent papers. **?** interviewed high school teachers on methods that they used to teach melodic dictation. **?** interviewed six sophomore music majors to find sucessful strategies that students enaged with when completing melodic dications. **?** reported beneficial effects to direct student's attention and guide them through melodic dictation exercises suggesting that some sort of mental organizaton of the dication process is helpful. **?** found that having students sing a preparatory singing pattern after hearing the target melody, essentially a distractor task, hindered performance on melodic dictation. **?**... **?** found no effects of test presentation format (visual versus aural-visual) using a melodic memory paradigm. **?** reported no significant advatage to listening strategies while partaking in a melodic dictation test.

#### 2.3.1.1 Recent Computational Musicology Work papers and findings

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

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

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

#### 2.3.1.2 Static Views of Computational Features/ FANTASTIC

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

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

Work using the FANTASTIC toolboox has been sucessful in predicting court case decisions (**?**), predicting chart sucesses of songs on the Beatles' *Revolver* (**?**), 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 (**??**) and has even been the basis of the development of a computer assistted 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 Informaton Dynamics Of Melody IDyOM models musical expectancy using various information theoretic concepts inspired by Claude Shannon (SHANNON). The model takes an unsupervised machine learning approach and calculates the information content of the amount of DECLARED n-grams in the corpus. As a model exploring expection for melody IDyOM has has been applied to a variety of settings LIST THEM HERE. The domain general application of IDyOM has given creedence to Meyer's assertion that the enculturation of musical styles stems from statistical exposure to melodies and be somewhat refelective of the cogitive processes used in muscical perception. IDyOM has also been recently extended to look at expectation in multi-part chorales (SAUVE WORK) and expectations of harmony (HARRISON WORK).

(?)

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

### 2.3.2 Experimental

- Advantage of vocal melodies paper | Weiss 2015
- Voice is better than instruments Weiss Trehub Schellenberg 2012
- Engaging, conspecific signal Weiss Peretz 2019 , not due to expressivity. even see this in amusics.
- Not affected by musical training or exposure
- MIR stuff
- Stuff out of education stuff that is also experimental
- WM W LVH, two span tasks and claim it effects langauge tasks
- WM Halpern 2017 with pearce, LTM stuff , re look this up

## 2.4 Modeling and Polymorphism of Ability

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

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

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

## 2.5 Conclusions

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

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

### 2.5.1 Add In

- (**?**)
- Dowling 1991 (**?**)

# Chapter 3

# Individual Differences

## 3.1 Rationale

The first two steps of Gary Karpinski's model of melodic dictation (**??**) rely exclusively on the mental representation of melodic information. Karpinski conceptualizes the first stage of *hearing* as involving the physical motions on the tympanic membrane, as well as the listener's attention to the musical stimulus. This stage is distinguished from that of *short-term melodic memory* which refers to the amount of melodic information that can be represented in conscious awareness. Given that neither stage of the first two steps of Karpinski's model requires any sort of musical expertise, every individual with normal hearing and cognition should be able to partake in the first two steps of melodic dictation.[1] The ability to hear, then remember musical information is where all students of melodic dictation are presumed to begin their aural skills education. From this baseline, students receive explicit education in music theory and aural skills to develop the ability to link they hear to what can then be musically understood and consequently notated.

While the majority of beginning students of melodic dictation are assumed to start at the same ability, cognitive psychology research suggests that individual differences in cognitive ability exist and must be accounted for from a psychological and pedagogical perspective (**??**)[2]. In order to fully capture the diversity of listening abilities among students of melodic dictation, a complete account of melodic dictation must include individual differences in ability. Understanding how differences at the individual level vary also will help pedagogues know what can be reasonably expected of students with different experiences and abilities.

Attempting to investigate all four parts of melodic dictation from hearing, to short-term melodic memory, to musical understanding, to notation is cumbersome both from a theoretical perspective and practically unfeasible due to the amount of variables that contribute to this process. In order to obtain a clearer picture of what mechanisms contribute to this process, these steps must be be investigated in turn. This chapter investigates the first two steps of the Karpinski model of melodic dictation (**??**) with an experiment examining individual factors that contribute to musical memory that do not depend on knowledge of Western musical notation. By understanding which, if any, individual factors play a role in this process, it will inform what can be reasonably expected of individuals when other musical variables are then introduced.

[1] This whole model needs critique under the WMC literature. It's kind of strange to think that the act of something hitting your ear is different than attention (the way that Cowan thinks about WMC and again that you can split up the representation in memory from that of what the characteristics are of the melody like the meter and scale degrees, which have been argued to be part of intrinsic qualia) also there is a big problem here about stuff being actively rehearsed or not

[2] Is there a better way to set this up?

## 3.2 Individual Differences

### 3.2.1 Improving Musical Memory

Most aural skills pedagogy assumes students begin with approximately the same baseline listening and dictation abilities. Assuming this baseline allows teachers to cover requisite information systematically and ensure that students are given the the same tools to enable their success in the classroom. This assumption of similar baseline of abilities is implicit in the Karpinski model of melodic dictation. The model provides a framework of mental choreography students are encouraged to build upon that is agnostic to individual differences; Karpinski's model assumes that all individuals regardless of their background will engage in the same process. As students gain more knowledge in music theory, they build their musical understanding which in turn enables them to recognize more of the auditory scene they are focusing on. In addition to learning explicit knowledge to facilitate their musical understanding, Karpinski suggests that there are two other skills that students can develop in order to improve their short-term musical memory: extractive listening and chunking. In Karpinski's own words:"Only one or both strategies can extend the capacity of short-term musical memory: (1) extractive listening and (2) chunking. (pp. 71)"

Karpinski defines extractive listening as "a combination of focused attention and selective memorization (p.70)". Extractive listening requires students to be able to focus on the material they will be mentally representing and tune out other sources of stimulation that might distract the student. In order to improve this ability, Karpinski suggests practicing listening to melodies and having students practice directing their attention to pre-determined set sequences of notes. Students should slowly work towards being able to auralize the melody with other musical information still sounding. Karpinski claims that honing one's attention via this type of progressive practice will not only improve student's ability to dictate melodies, but also help them with a host of other musical activity. Further, Timothy Chenette has since proposed similar types of progressive loading aural exercises by co-opting standard cognitive tasks used in working memory paradigms (**?**) in order to help students improve their ability to focus in aural skills.

After students master the ability to selectively hear and retain portion of a melody, the other way in which they can improve their dictation abilities is via chunking. Chunking is a listener's ability group smaller units of musical material into a larger group. The idea of chunking derives from earlier work from Gestalt psychologists and was one of the initial mechanisms proposed by **?** able to extend the finite window of memory. The general idea is that if a collection of notes can be identified as its own discrete entity– such as a descending major triad in first inversion– the listener will only have to remember that one structure, rather than its component parts. As discussed in the previous chapter in Parallels Between Working Memory and Melodic Dictation, music's inherently sequential nature affords it many opportunities to find repeated patterns which can be labeled, musically understood, and thus chunked. While stimuli that are inherently sequential are problematic for psychologists investigating capacity limits of working memory capacity (**?**), students are expected to use chunking to their advantage in order to become more adept listeners. As students learn to chunk more efficiently, they are able to process more musical information in their short-term musical memory. With the development of both skills, students are presumed to increase their musical memory and ultimately improve their melodic dictation abilities. But what evidence supports the assertion that individuals are able to improve on their ability to both learn and remember melodies?

### 3.2.2 Memory for Melodies

Research findings from the memory for melody literature are mixed when considering how people vary in their ability to remember musical material (**?**). For example, no effect of an individual's musical training was found by **?** in a paradigm where both musically trained and non-musically trained individuals were presented with melodies using a recognition paradigm task with melodies over the course of two days. In a musical recognition task, **?** found no effect of musicianship on memory. Using a recognition paradigm, **?** found an effect of musical training on melodic memory, but the significant effect reported was not found in correctly identifying melodies, but rather in correctly identifying melodies that they had not heard before. **?** reported no effects of musical training on their recognition paradigm experiment. They however did not include any expert participants in their sample and the focus of this particular study was to look at structural features of the melody, rather than individual level features. Additionally, other studies have also found that musical expertise is not a successful predictor of melodic recognition (**??**). As with much of the music psychology literature, one of the reasons that these studies may have not found a memory advantage for the more musically trained is that how musical training is measured varies widely from study to study (**?**). This inability to measure musical exposure additionally complicates controlling for the amount of variability of what might drive the memory effects in the models of musical memory. When measured continuously using paradigms that require immediate recall and judgment, musical training does often predict memory for musical materials.

Using a step wise modeling procedure, **?** consistently found evidence that musical training to be a significant predictor of ability to preform well on a melodic discrimination task when developing an item response theory based test of melodic memory. Using regression modeling, Harrison et. al reported to be able to explain a large amount of the variance ($R^2 = 0.459$) when reporting response variability in a melodic discrimination task (**?**) when measuring musical training via the Goldsmith's Musical Sophistication Index (**?**). **?** found musical training, when measured continuously, was able to be a significant predictor using a exposure-recall paradigm among other predictor variables.

Even despite mixed evidence suggesting different effects of musical training on an individual's ability to remember melodies, it is important to note that these studies do not specifically deal with melodic dictation, and thus cannot be used as a perfect comparison for a number of reasons. The first is that melodic dictation is a much more complicated process that not only involves hearing a melody after a few iterations, but also its notation. Seeing as students need to notate their melodies, which again is dependent on their knowledge of Western musical notation, melodic dictation is secondly a more cognitively demanding process than the previously mentioned studies on memory for melody which often only require a simple discrimination.

### 3.2.3 Musician's Cognitive Advantage

While the above memory for melodies literature is mixed regarding the musician's advantage, there is research from cognitive psychology to support the latter evidence of an advantage of musical training in perceptual tests. Some researchers suggest that musicians have better cognitive abilities on a more domain general level, which could lead to better performance and explain differences in performance. Work as reviewed in **?** investigating the relationship between musical training and general intelligence suggest that both children and adults who engage in musical activity tend to score higher on general measures of intelligence than their non-musical peers (**????**). Importantly, this association between intelligence and musical training comes with a correlation between duration of musical training and the extent of the increases in intelligence (**????**). While many of these studies are correlation, other researchers have further investigated this relationship in experimental settings in attempt to control for confounding variables like socio-economic status and parental involvement in out of school activities (**?????**), but findings have been mixed.

Schellenberg (**?**) notes that in many of these studies there is a problem of too small of a sample size in his review (**???**) in that studies that are typically smaller might be under powered to detect any effects. Also referenced in Schellenberg's review is evidence that when professional musicians are matched with non-musicians from the general population these associations are non-existent (**?**). Interpreting the current literature Schellenberg puts forward the hypothesis that higher functioning children might self-select into music lessons and they tend to stay in lessons longer which leads to the observed differences in intelligence. Additionally, Schellenberg remains skeptical of any sorts of causal factors regarding increases in IQ (**??**) noting methodological problems such as short exposure times or researchers who did not holding pre-existing cognitive abilities constant (**?**).

In addition to general intelligence, another cognitive ability where musicians tend to exhibit superior performance is that of memory. **?**'s meta-analysis investigating musical training and memory found not only a general advantage of musicians, but noted that musicians tended to perform better on memory tasks especially in cases where stimuli were short and tonal. This musician advantage could derive from a musician's ability to chunk information more effectively based on past exposure via implicit learning practices

(??). This difference also might reflect the above mentioned self-selection of higher functioning individuals to partake in music, which then explain the differences in memory.

As noted above, much of the research at this point still very much focuses on higher level relationships, which is progressively being improved upon by agreeing on how to measure what is actually driving these effects. Until more concrete theories emerge that link specific musical traits to music ability, music psychology will not be able to put forward clearer models of causal effects (?).

### 3.2.4 Relationship Established

Regardless of the direction of causality, the evidence discussed suggests that there is a relationship between musical training and cognitive ability. Clearly cognitive ability is at play in many tasks of perception and production and presumably these abilities would interact with other variables of interest such as musical training as theorized by some researchers above. Even in studies outside of music, domain general cognitive abilities have been shown to be predictive above and beyond domain specific expertise. In reviewing the current literature, (?) reiterate that while there is evidence some of the time in many domain specific areas like chess, games, and music, the current state of the literature is not definitive enough to explain exactly how this phenomena works on a global level.

Though of all the studies mentioned thus far, one cognitive ability deserving of special attention is that of working memory. As noted by (?), many tests of memory– such as the tests above–require the encoding and active manipulation of musical material. In his 1995 article, Berz draws important parallels between working memory systems and music tests and postulated new loop.

For example ? found working memory to be predictive of performance in a sight reading task above and beyond that of deliberate practice. Work by Kopiez (??) has additionally linked the importance of working memory to performance on sight reading tasks. In multiple studies, Andrea Halpern and colleagues have also shown measures of working memory to be linked to performance in musical production tasks (??) and has even interpreted these findings in terms of Berz's memory loop. Other work by ? has also made important links to an individual's ability to remember and recall musical information and working memory. Harrison and colleagues put forward a cognitive model based on research in working memory that predicted which features of a melody– based on theoretical considerations from working memory– would be best at predicting behavioral performance. They proposed that perceptual encoding, memory retention, similarity comparison, and decision-making could be used to contextualize differences in their memory recognition paradigm. While they did find evidence to support these notions, they did not take any domain general measures of working memory capacity and thus were unable to conclude if domain general processes were able to better explain their data than using individual level predictors.

Additionally, ? used a latent variable approach where they investigated executive function in a sample of 161 university students. Using Miyake's conception of executive function (??) and mixed effects modeling, Okada and Slevc found an effect of musical training as measures withe Goldsmiths Musical Sophistication Index on the updating component of the executive functioning model, a construct often interpreted as similar to working memory capacity. Okada and Slevc did not however link performance on their executive functioning tasks to an objective measure of musical performance implemented by the Goldsmiths Musical Sophistication Index.

### 3.2.5 Dictation Without Dictation

So given the complex network of variables at play, in order to understand how these individual factors affect the first two steps of melodic dictation, a multivariate approach is needed. In order to investigate the effects of individual factors on baseline, I must first assume that using a melodic discrimination paradigm can be used as a proxy for the first two steps of the Karpinski model of melodic dictation. I argue that because melodic discrimination paradigms require perceptual encoding, memory retention, and two other cognitive manipulations of similarity comparison and decision making as argued by ?, this paradigms do in fact resemble the first two steps of the Karpinski model. Karpinski's hearing and short-term musical memory could just as easily be described as perceptual encoding and memory retention. Additionally, the requirement to execute a decision while representing musical information in memory– Harrison and colleague's Similarity Comparison and Descion making– can be mapped on to later stages of Karpinski's model of musical understanding, and subsequently notation.

One of the most complete suites of measuring musicality that employs both objective and subjective measures of musical sophistication is the Goldsmiths Musical Sophistication Index or Gold-MSI (**?**). The Gold-MSI has both a self-report questionnaire as well as two test of objective ability. One of the tests employs a beat detection paradigm, the other is a melodic discrimination paradigm. Seeing as both measures mirror tasks used in the aural skills classroom and the two are purported to measure different constructs, both will be used in this study. Since its initial publication, adaptive short forms of the tests have been developed using item response theory (**?**). These tests were not available to be used at the time of this study's data collection.

Assuming that a melodic discrimination task can then stand in for the first two steps of the Karpinski model, I can then model the relationships between performance on this musical memory task with individual level variables using structural equation modeling. By doing this I can examine the extent to which, if any, factors contribute to the first two steps of melodic dictation.

### 3.2.6 Cognitive Measures of Interest

Having previously established that many tests of musical ability and aptitude may in fact be tests of working memory (**?**), one factor not yet accounted for in the memory for melodies literature is a domain general measure of working memory. If working memory is conceptualized using Cowan's model of working memory as the window of attention (**?**), measuring working memory would need to be operationalized using a task that implements both the retention and manipulation of information in memory. This is commonly done with complex span tasks (**?**). Complex span tasks, unlike simple span tasks like the *n-back* paradigms, require both the retention and manipulation of items in memory and thus better reflect a Cowanian model of working memory (**?**).

Additionally, since general intelligence is often predictive of performance on a host of cognitive tasks such as educational success, income, and even life expectancy (**?**) and has been theoretically related to working memory (**?**), this measure should also be accounted for when investigating individual features that contribute to the first two steps of melodic dictation using a standard paradigms of intelligence testing (**??**). Finally, in response to claims made by **?**, having to need to account for specific covariates, this study also will track socioeconomic status and degree of education, variables used in previous music psychology research (**??**).

### 3.2.7 Structural Equation Modeling

Given the complex nature being investigated and the theoretical concepts at play such as working memory, general fluid intelligence, and musical sophistication conceptualized as a latent variable, it follows that the most appropriate method of parsing out the variance in this covariance structure would be to use some form of structural equation modeling (**?**). Structural equation modeling uses latent variables, theoretical constructs thought to exist yet are not possible to measure directly, that using a closed set of algebraic systems originally developed by Sewall Wright (**?**). When used under the right conditions, the technique is powerful enough to determine casual mechanisms in closed systems (**?**), but this is not the case in this analysis.

### 3.2.8 Hypotheses

If I then assume that a same-different melodic memory paradigm is a stable proxy for the first two steps of Karpinski's model of melodic dictation, then data generated from both objective tests of the Goldsmiths' Musical Sophistication Index can serve as proxy for this measure of interest. In this analyses, I will use a series of structural equation models in order to investigate how various individual factors contribute to an individual's memory for melody. Following a step-wise procedure, these sets of analyses will provide a way to investigate what individual factors need to be accounted for in future research.

Given a robust instrument for measuring musicality, and two well established cognitive measures as specifically defined below, this analysis seeks to investigate the degree to which these individual level variables are predictive of a task that is proxy to the first two steps of melodic dictation. If a large proportion of the variance of musical memory can be attributed to training, then variables related to the Goldsmiths Musical Sophistication Index should be most predictive with the highest path coefficients and lead to the best model fit. If instead cognitive factors do play a role, this should be evident in the path loadings.

## 3.3 Overview of Experiment

### 3.3.1 Participants

Two hundred fifty-four students enrolled at Louisiana State University completed the study. Students were mainly recruited in the Department of Psychology and the School of Music. The criteria for inclusion in the analysis were no self-reported hearing loss, not actively taking medication that would alter cognitive performance, and the removal of any univariate outliers (defined as individuals whose performance on any task was greater than 3 standard deviations from the mean score of that task). Using these criteria, eight participants were not eligible due to self reporting hearing loss, one participant was removed for age, and six participants were eliminated as univariate outliers due to performance on one or more of the tasks of working memory capacity. Thus, 239 participants met the criteria for inclusion. The eligible participants were between the ages of 17 and 43 (M = 19.72, SD = 2.74; 148 females). Participants volunteered, received course credit, or were paid $20.

### 3.3.2 Materials

#### 3.3.2.1 Cognitive Measures

All variables used for modeling approximated normal distributions. Processing errors for each task were positively skewed for the complex span tasks similar to **?**. Positive and significant correlations were found between recall scores on the three tasks measuring working memory capacity (WMC) and the two measuring general fluid intelligence (Gf). The WMC recall scores negatively correlated with the reported number of errors in each task, suggesting that rehearsal processes were effectively limited by the processing tasks (**?**).

#### 3.3.2.2 Measures

##### 3.3.2.2.1 Goldsmiths Musical Sophistication Index Self Report (Gold-MSI)

Participants completed a 38-item self-report inventory and questions consisted of free response answers or choosing a selection on a likert scale that ranged from 1-7. (**?**). The complete survey with all questions used can be found at goo.gl/dqtSaB.

##### 3.3.2.2.2 Tone Span (TSPAN)

Participants completed a two-step math operation and then tried to remember three different tones in an alternating sequence (based upon **?**). We modeled the three tones after **?** paper's using frequencies outside of the equal tempered system (200Hz, 375Hz, 702Hz). The same math operation procedure as OSPAN was used. The tones was presented aurally for 1000ms after each math operation. During tone recall, participants were presented three different options H M and L (High, Medium, and Low), each with its own check box.

Tones were recalled in serial order by clicking on each tone's box in the appropriate order. Tone recall was untimed. Participants were provided practice trials and similar to OSPAN, the test procedure included three trials of each list length (3-7 tones), totaling 75 letters and 75 math operations.

#### 3.3.2.2.3 Operation Span (OSPAN)

Participants completed a two-step math operation and then tried to remember a letter (F, H, J, K, L, N, P, Q, R, S, T, or Y) in an alternating sequence (**?**). The same math operation procedure as TSPAN was used. The letter was presented visually for 1000ms after each math operation. During letter recall, participants saw a 4 x 3 matrix of all possible letters, each with its own check box. Letters were recalled in serial order by clicking on each letter's box in the appropriate order. Letter recall was untimed. Participants were provided practice trials and similar to TSPAN, the test procedure included three trials of each list length (3-7 letters), totalling 75 letters and 75 math operations.

#### 3.3.2.2.4 Symmetry Span (SSPAN)

Participants completed a two-step symmetry judgment and were prompted to recall a visually-presented red square on a 4 X 4 matrix (**?**). In the symmetry judgment, participants were shown an 8 x 8 matrix with random squares filled in blank. Participants had to decide if the black squares were symmetrical about the matrix's vertical axis and then click the screen. Next, they were shown a "yes" and "no" box and clicked on the appropriate box. Participants then saw a 4 X 4 matrix for 650 ms with one red square after each symmetry judgment. During square recall, participants recalled the location of each red square by clicking on the appropriate cell in serial order. Participants were provided practice trials to become familiar with the procedure. The test procedure included three trials of each list length (2-5 red squares), totalling 42 squares and 42 symmetry judgments.

#### 3.3.2.2.5 Gold-MSI Beat Perception

Participants were presented 18 excerpts of instrumental music from rock, jazz, and classical genres (**?**). Each excerpt was presented for 10 to 16s through headphones and had a tempo ranging from 86 to 165 beats per minute. A metronome beep was played over each excerpt either on or off the beat. Half of the excerpts had a beep on the beat, and the other half had a beep off the beat. After each excerpt was played, participants answered if the metronome beep was on or off the beat and provided their confidence: "I am sure", "I am somewhat sure", or "I am guessing". The final score was the proportion of correct responses on the beat judgment.

#### 3.3.2.2.6 Gold-MSI Melodic Memory Test

Participants were presented melodies between 10 to 17 notes long through headphones (**?**). There were 12 trials, half with the same melody and half with different melodies. During each trial, two versions of a melody were presented. The second version was transposed to a different key. In half of the second version melodies, a note was changed a step up or down from its original position in the structure of the melody. After each trial, participants answered if the two melodies had identical pitch interval structures.

#### 3.3.2.2.7 Number Series

Participants were presented with a series of numbers with an underlying pattern. After being given two example problems to solve, participants had 4.5 minutes in order to solve 15 different problems. Each trial had 5 different options as possible answers (**?**).

#### 3.3.2.2.8 Raven's Advanced Progressive Matrices

Participants were presented a 3 x 3 matrix of geometric patterns with one pattern missing (?). Up to eight pattern choices were given at the bottom of the screen. Participants had to click the choice that correctly fit the pattern above. There were three blocks of 12 problems, totalling 36 problems. The items increased in difficulty across each block. A maximum of 5 min was allotted for each block, totalling 15 min. The final score was the total number of correct responses across the three blocks.

## 3.3.3 Procedure

Participants in this experiment completed eight different tasks, lasting about 90 minutes in duration. The tasks consisted of the Gold-MSI self-report inventory, coupled with the Short Test of Musical Preferences, and a supplementary demographic questionnaire that included questions about socioeconomic status, aural skills history, hearing loss, and any medication that might affect their ability to perform on cognitive tests. Following the survey they completed three WMC tasks: a novel Tonal Span, Symmetry span, and Operation span task; a battery of perceptual tests from the Gold-MSI (Melodic Memory, Beat Perception, Sound Similarity) and two tests of general fluid intelligence (Gf): Number Series and Raven's Advanced Progressive Matrices.

Each task was administered in the order listed above on a desktop computer. Sounds were presented at a comfortable listening level for the tasks that required headphones. All participants provided informed consent and were debriefed. Only measures used in modeling are reported below.

## 3.3.4 Results

### 3.3.4.1 Descriptive, Data Screening, Correlational

The goal of the analyses was to examine the relationships among the measures and constructs of WMC, general fluid intelligence, musical sophistication (operationalized as the General score from the Gold-MSI), in relation to the two objective listening tests on the Gold-MSI. Before running any sort of modeling, we inspected our data to ensure in addition to outlier issues as mentioned above, the data exhibited normal distributions. I report both correlation values, as well as visually displaying our distributions in Figure 1.

Before running any modeling, I checked our data for assumptions of normality since violations of normality can strongly affect the covariances between items. While some items in Figure 1 displayed a negative skew, many of the individual level items from the self report scale exhibited high levels of Skew and Kurtosis beyond the generally accepted ± 2 (?), but none of the items with the unsatisfactory measures are used in the general factor.

### 3.3.4.2 Modeling

#### 3.3.4.2.1 Measurement Model

I then fit a measurement model to examine the underlying structure of the variables of interest used to assess the latent constructs (general musical sophistication, WMC, general fluid intelligence) by performing a confirmatory factor analysis (CFA) using the lavaan package (?) using R (?). Model fits in can be found in Table X. For each model, latent factors were constrained to have a mean of 0 and variance of 1 in order to allow the latent covariances to be interpreted as correlations. Since the objective measures were on different scales, all variables were converted to z scores before running any modeling.

Variables are listed in the table below:

| Abbreviation | Variable |
|---|---|
| gen | General Self-Report Musical Sophistication |

| Abbreviation | Variable |
|---|---|
| wmc | Working Memory Capacity |
| gf | General Fluid Intelligence |
| zIS | Identify What is Special |
| zHO | Hear Once Sing Back |
| zSB | Sing Back After 2-3 |
| zDS | Don't Sing In Public |
| zSH | Sing In Harmony |
| zJI | Join In |
| zNI | Number of Instruments |
| zRP | Regular Practice |
| zNCS | Not Consider Self Musician |
| zNcV | Never Complimented |
| zST | Self Tonal |
| zCP | Compare Performances |
| zAd | Addiction |
| zSI | Search Internet |
| zWr | Writing About Music |
| zFr | Free Time |
| zTP | Tone Span |
| zMS | Symmetry Span |
| zMO | Operation Span |
| zRA | Ravens |
| zAN | Number Series |

#### 3.3.4.3 Structural Equation Models

Following the initial measurement model, I then fit a series of structural equation models in order to investigate both the degree to which factor loadings changed when variables were removed from the model as well as the model fits. I began with a model incorporating our three latent variables (general musical sophistication, WMC, general fluid intelligence) predicting our two objective measures (beat perception and melodic memory scores) and then detailed steps we took in order to improve model fit. For each model, I calculated four model fits: $\chi^2$ , comparative fit index (CFI), root mean square error (RMSEA), and Tucker Lewis Index (TLI). In general, a non-significant $\chi^2$ indicates good model fit, but is overly sensitive to sample size. Comparative Fit Index (CFI) values of .95 or higher are considered to be indicative of good model fits as well as Root Mean Square Error (RMSEA) values of .06 or lower, Tucker Lewis Index (TLI) values closer to 1 indicate a better fit (**?**).

After running the first model (Model 1), I then examined the residuals between the correlation matrix the model expects and our actual correlation matrix looking for residuals above .1. While some variables scored near .1, two items dealing with being able to sing ("I can hear a melody once and sing it back after hearing it 2 – 3 times" and "I can hear a melody once and sing it back") exhibited a high level of correlation amongst the residuals (.41) and were removed for Model 2 and model fit improved significantly ($\chi^2$ (41)=123.39, $p < .001$).

After removing the poorly fitting items, I then proceeded to examine if removing the general musical sophistication self-report measures would significantly improve model fit for Model 3. Fit measures for Model 3 can be seen in Table 3 and removing the self-report items resulted in a significantly better model fit ( 2 (171)=438.8, $p < .001$). Following the rule of thumb that at least 3 variables should be used to define any latent-variable (**?**) I modeled WMC as latent variable and Gf as a composite average of the two tasks administered in order to improve model fit. This model resulted in significant improvement to the model ($\chi^2$ (4)=14.37, $p < .001$). Finally I examined the change in test statistics between Model 2 and a model that removed the cognitive measures– a model akin to one of the original models reported in (**?**)– for Model

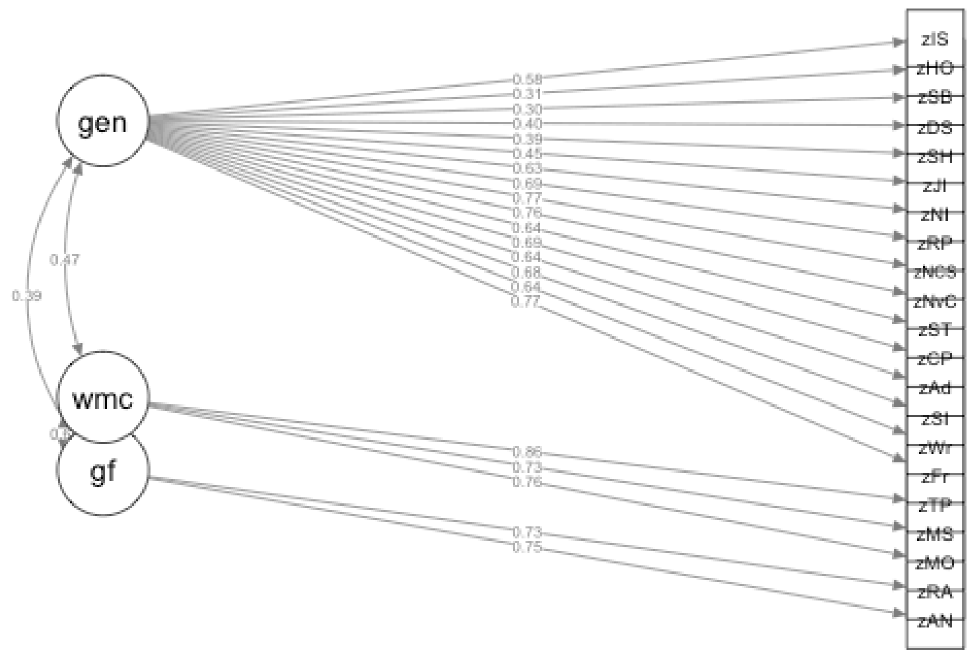


Figure 3.1: CFA Measurement Model

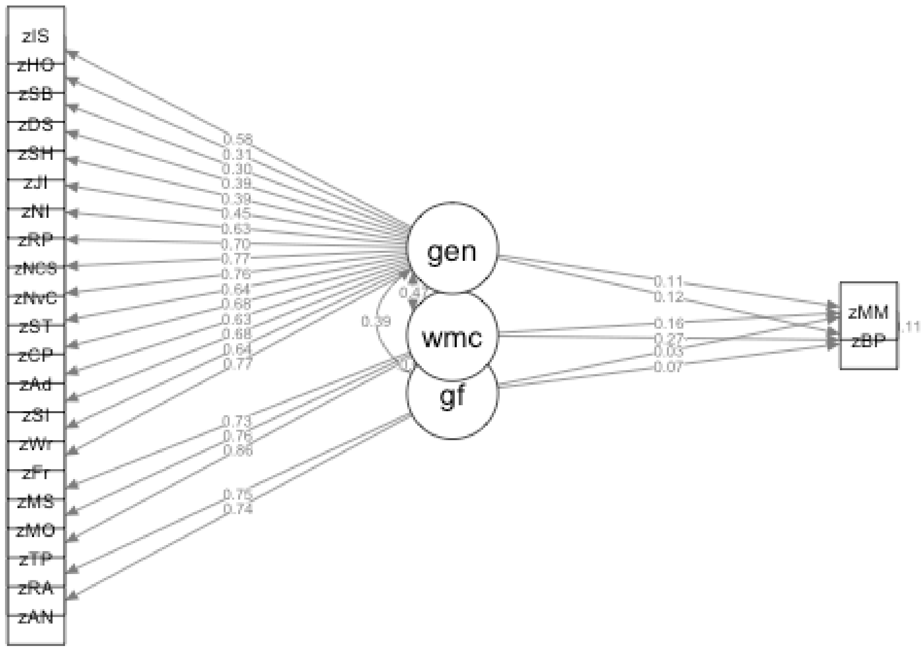


Figure 3.2: Full Model, All Variables Included

5. Testing between the two models resulted in a significant improvement in model fit ($\chi^2$ (78)=104.75, p < . 001). Figure X displays Model 4, our nested model with the best fit indices.

| Models | df | chi | p | CFI | RMSEA | TLI |
|---|---|---|---|---|---|---|
| CFA | 186 | 533.60 | > .001 | 0.83 | 0.09 | 0.81 |
| Model 1 | 222 | 586.30 | > .001 | 0.83 | 0.08 | 0.80 |
| Model 2 | 181 | 462.90 | > .001 | 0.86 | 0.08 | 0.83 |
| Model 3 | 10 | 24.11 | > .05 | 0.97 | 0.08 | 0.94 |
| Model 4 | 6 | 9.74 | > .14 | 0.99 | 0.51 | 0.97 |
| Model 5 | 130 | 358.16 | > .001 | 0.83 | 0.10 | 0.80 |

## 3.4 Discussion

### 3.4.1 Model Fits

#### 3.4.1.1 Measurement Model

After running a confirmatory factor analysis on the variables of interest, the model fit was below the threshold of what is considered a "good model fit" as shown in @ref(Model Fits) with references to above model fits. This finding is to be expected since no clear theoretical model has been put forward that would suggest that the general musical sophistication score, when modeled with two cognitive measures should have a "good" model fit. This model was run to create a baseline measurement.

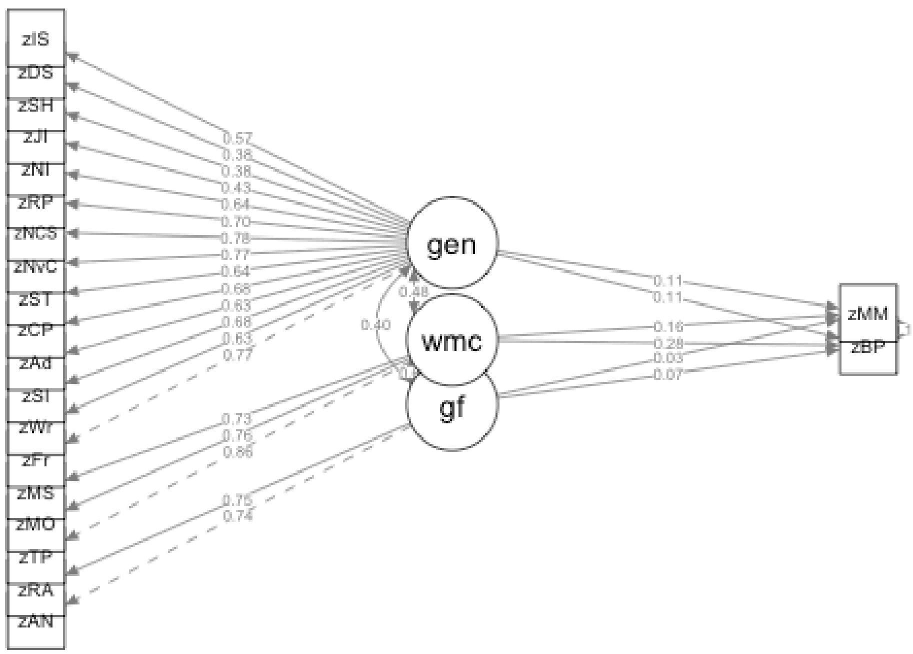


Figure 3.3: Full Model, Highly Correlated Residual Items

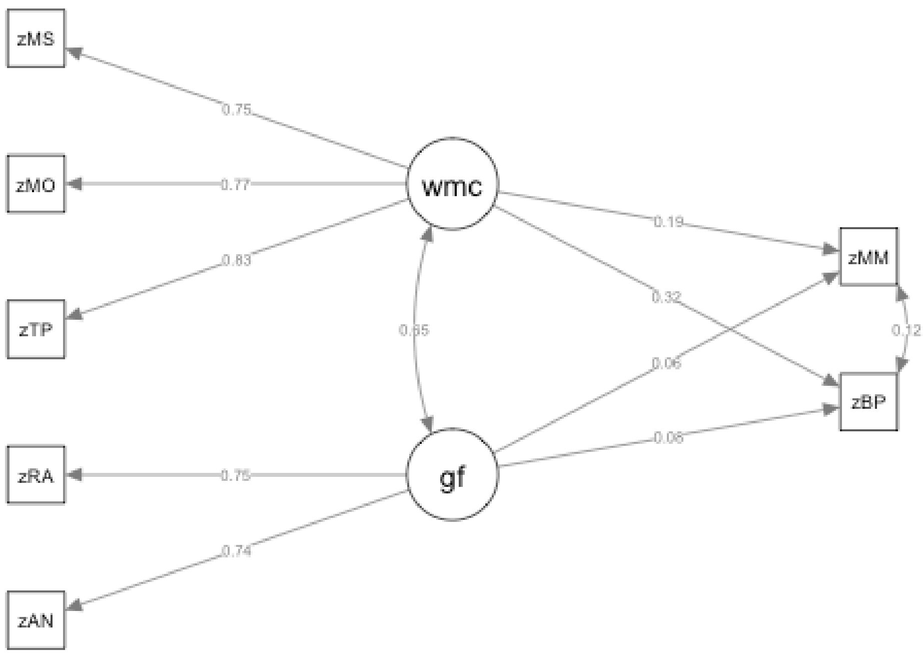


Figure 3.4: Self Report Removed, Only Cognitive Measures

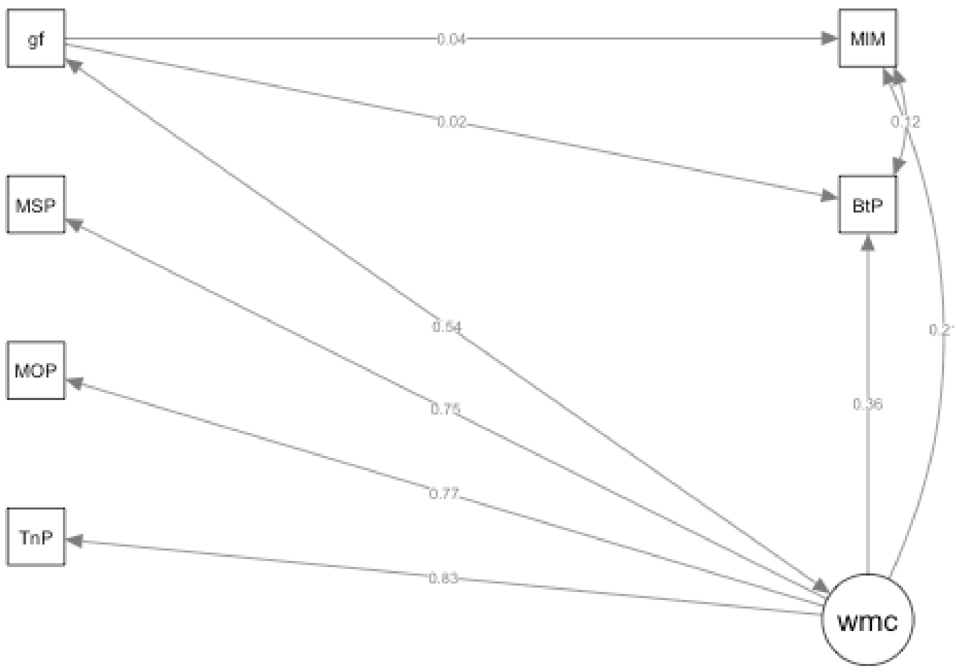


Figure 3.5: Cognitive Measures, Gf as Observed

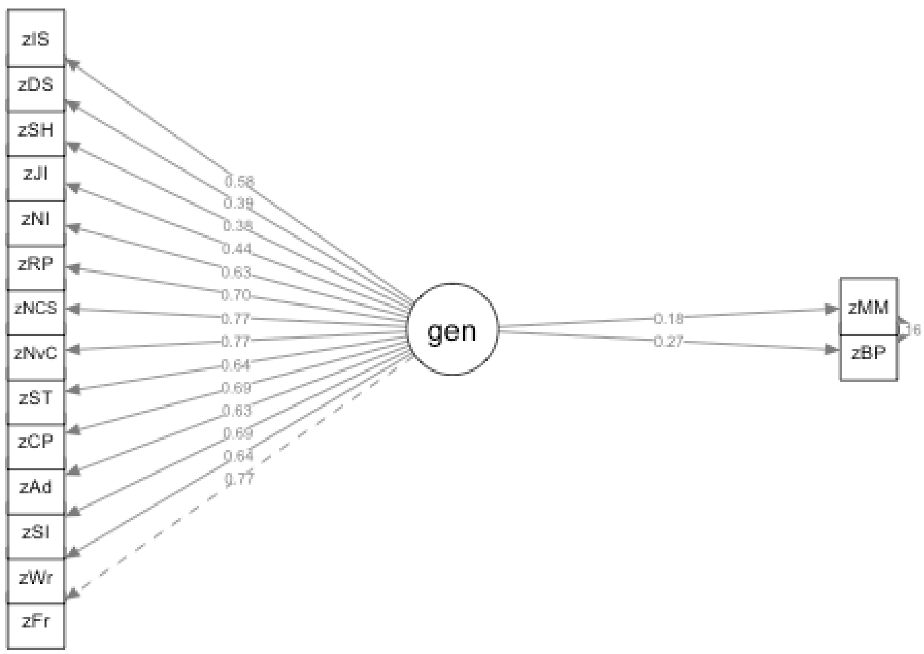


Figure 3.6: General Self Report Only

#### 3.4.1.2 Structural Equation Model Fitting

Following a series of nested model fits, we were able to improve model fits on a series of structural equation models that incorporated both measures of working memory capacity and measures of general fluid intelligence. Before commenting on new models, it is worth noting that the Model 5 does not seem to align with the findings from the original 2014 paper by (?). While the correlation between the objective tasks is the same (r = .16), the factor loadings from this paper suggest lower values for both Beat Perception (.37 original, .27 this paper) as well as Melodic Memory (.28 original, .18 this paper). Note that two items were removed dealing with melody for memory for this model; when those items were re-run with the data, the factor loadings did not deviate from these numbers.

The first two models I ran resulted in minor improvements to model fit. While difference in models was significant ($\chi^2$ (41)=123.39, p < . 001), probably due to the number of parameters that were now not constrained, the relative fit indices of the models did not change drastically. It was not until the self-report measures were removed from the model, and then manipulated according to latent variable modeling recommendations, was there a marked increase in the relative fit indices. Fitting the model with only the cognitive measures, I was able to enter the bounds of acceptable relative fit indices that were noted above. In order to finding evidence that the cognitive models (Models 3 and 4) were indeed a better fit than using the General factor, I additionally ran a comparison between our adjusted measurement model and a model with only the self-report. While both of the nested models were significantly different, the cognitive models exhibited superior relative fit indices. Lastly, turning to Figure 3, I note that the latent variable of working memory capacity exhibited much larger factor loadings predicting the two objective, perceptual tests than our measure of general fluid intelligence. I also note that the factor loading predicting the Beat Perception task (.36) was higher than that of the Melodic Memory task (.21). These rankings mirror that of the original (?) paper and merit further examination in order to disentangle what processes are contributing to both tasks.

Given the results here that suggest that measures of cognitive ability play a significant role in tasks of musical perception, we suggest that future research should consider taking measures of cognitive ability into account, so that other variables of interest are able to be shown to contribute above and beyond baseline cognitive measures.

### 3.4.2 Relating to Melodic Dictation

This study sought to investigate the extent to which individual factors contributed to an individual's ability to perform the first two steps of melodic dictation. In order to do this, I assumed that the first two steps of the Karpinski model– hearing and short term melodic memory– could be investigated by using a same-different melodic memory paradigm. Both task require the dual activation of representing information in conscious awareness and completing a cognitive task. Using this paradigm also allowed me to investigate the first to steps of Karpinski's model using both individuals with and without musical training.

Overall, when interpreting the results I found evidence to corroborate claims made by ? positing the importance of working memory in both tests of musical aptitude, and consequently the first two steps of melodic dictation as described by Karpinski. Relatively, working memory seemed to dominate as the variable with the most explanatory power as derived from both the best overall model fits and highest path coefficients in the latent variable modeling. This is not a surprising finding given the context, yet has major implications for future research in music perception. If a domain general process is able to predict performance on a domain specific task (melodic memory) better than measures of self report and training, future studies in music perception will need to be able to demonstrate how the process they purport to be the driving factor behind their models explains their findings above and beyond working memory capacity.

Also worth discussing is why general fluid intelligence did not fare as well in the models above. One reason that this might be is because general intelligence tests are designed in two ways differing from that of melodic dictation. The first is that general fluid intelligence tests administered here do not have any time component to them. While tasks like Raven's matrices (?) and the number series (?) tests are timed, the information is presented visually to participants. The second is that general fluid intelligence is designed to measure abilities outside of the context of previously known information (**?**) and questions surrounding music perception depend both principles of statistical learning (**???**) and stylistic enculturation (**???**). General fluid intelligence might be helpful at later stages of cognitive processing such as the musical understanding and notation phases of the Karpinski model, but their effect does not seem to be present here.

From a pedagogical standpoint, this is important in that many teachers are aware that students will vary in terms of their working memory ability. While it would be statistically rare to actually find someone with a working memory deficit, knowing that this construct is powerful predictor of performance at such an early stage of melodic dictation reinforces that teachers should be aware of it. One practical consideration for the classroom within the Karpinksi framework would be to encourage students to listen for smaller chunks when using extractive listening. Using a Cowanian model of working memory, students should extract smaller chunks so that they still have cognitive resources available in order to focus on the later stages of the Karpinski model (musical understanding and notation). As attention is limited, not listening to more than you can hold will free up cognitive resources that might later be used in melodic dictation. Further students could take up recommendations like that of **?** and focus on activities that might help them increase their ability to focus, knowing that this practice will most likely not increase their working memory.

Not only will this finding have relevance in the classroom, but this findings suggests that future work looking to do more robust modeling of melodic dictation must take into account the window of attention. In chapter X, I incorporate this finding into a computational model of melodic dictation and use the finite window of working memory as a perceptual bottleneck to constrain incoming musical information.

In this chapter I fit a series of structural equation models in order to investigate the degree to which baseline cognitive ability was able to predict performance on a musical perception task. My findings suggest that measures of working memory capacity are able to account for a large amount of variance beyond that of self report in tasks of musical perception.

# Chapter 4

# Computation Chapter

## 4.1 Rationale

Music theorists use their experience and intuitions to build appropriate curricula for their aural skills pedagogy. Teaching aural skills typically starts with providing students with simpler exercises, often employing a limited number of notes and rhythms, and then slowly progressing to more difficult repertoire. This progression from simpler to more difficult exercises is evident in aural skills text books. Of the major aural skills textbooks such as the **?**, **?**, **?**, and **?**, each is structured in a way that musical material presented earlier in the book is more manageable than that nearer the end. In fact, this is true of almost any étude book: open to a random page in a book of musical studies and the difficulty of the study will likely scale accordingly to its relative position in the textbook. But it is not a melody's position in a textbook that makes it difficult to perform: this difficulty comes from the structural elements from the music itself.

Intuitively, music theorists have a general understanding of what makes a melody difficult to dictate. Factors that might contribute to this complexity could range from the number of notes in the melody, to the intricacies of the rhythms involved, to the scale from which the melody derives, or even more intuitively understood factors such as how tonal the melody sounds. Although given all these factors, there is no definitive combination of features that perfectly predicts the degree to which pedagogues will agree how complex a melody is. In many ways, questions of melodic complexity are very much like questions of melodic similarity: it depends on both who is asking the question and for what reasons (**?**).

Looking at the melodies presented in Figures X and Y, most aural skills pedagogues will be able to successfully intuit which melody is more complex, and presumably, more difficulty to dictate.

- FIGURE 1 – Melody with 8 Bars, functional accidentals (V/V, V/IV)
- FIGURE 2 – Same sets of notes rearranged

While I reserve an extended discussion of what features might characterize why one melody is more difficult to dictate than the other for this chapter, I assume that these melodies differ in their ability to be dictated in some fundamental way when performed in a similar fashion. Additionally, many readers of this dissertation can draw from anecdotal evidence of their own as to how students at various stages of their aural training might fair when asked to dictate both melodies. For some, melody Y might be overwhelmingly difficult.

In fact, melody Y might be overwhelmingly difficult for the vast majority of musicians to dictate. From a pedagogical standpoint, educators need to be able to know how difficult melodies are to dictate in order to ensure a degree of fairness when assessing a student's performance. While of course with each student there are inevitably many variables at play in aural skills instruction ranging from personal abilities, to the goals of the instructor in the scope of their course, I find it fair to claim that pedagogues assume that students will be expected to pass pre-established benchmarks throughout their aural skills education. As students progress, they are expected to be able to dictate more difficult melodies, yet exactly what makes a melody complex and thus difficult to dictate is often left to the expertise and intuition of a pedagogue. Intuition is an important skill for teachers to cultivate, but when it comes to determining objective measures of judgment, reserach from decison making science tends to suggest that no matter the expertiese, collective and objective knowledge tends to outperform a single person's judgment (KAHHMEAN AND TVERSKY, HEART ATTACK PAPER, LOGISITC REGRESSION THING). Having more clearly defined performance benchmarks also helps remove biases in grading that teachers may or may not be explicitly aware of.

In this chapter I survey and examine how tools from computational musicology can be used to help model an aural skills pedagogue's notion of complexity in melodies. First, I establish that theorists agree on the differences in melodic complexity using results from a survey of 40 aural skills pedagogues. Second, I explore how both static and dynamic computationally derived abstracted features of melodies can and cannot be used to approximate an aural skills pedagogue's intuition. Third and finally, I use evidence afforded by research in computational musicology to posit that the distributional patterns in a corpus of music can be strategically employed to create a more linear path to success among students of aural skills. I demonstrate how combining evidence from the statistical learning hypothesis, the probabilistic prediction hypothesis, and a newly posited distributional frequency hypothesis, it is possible to explain why some musical sequences in a melody are easier to dictate than others. Using this logic, I then create a new compendium of melodic incipits, sorted by their perceptual complexity, that can be used for teaching applications.

## 4.2 Agreeing on Complexity

Returning to melodies X and Y from above, an aural skills pedagogue most likely has an intuition to which of the two melodies X or Y would be easier to dictate. Melody X exhibits a predictable melodic syntax and phrase structure, the chromatic notes resolve within the conventions of the Common Practice period, and many of the melodic motives outline and imply a harmony based on tertian harmony. On the other hand, Melody Y's syntax does not conform to the conventions of the Common Practice period and does not imply any sort of underlying harmony. The duration of the rhythms appear irregular and the melody implies an uneven phrase structure. Yet both melodies X and Y have the exact same set of notes and rhythms. Though despite these content similarities, it would be safe to assume that melody X is probably much easier to dictate than melody Y assuming both were to be played in a similar fashion.

In fact, aural skills pedagogues tend to agree for the most part on questions of difficulty of dictation. To demonstrate this, I surveyed 40 aural skills pedagogues who all have taught aural skills at the post-secondary level. In this survey, participants were asked the questions presented in TABLE X and TABLE Y using a sample of 20 melodies found in the a commonly used sight-singing text book (**?**). I present the details of the survey below.

### 4.2.1 Methods

To select the melodies used in this survey, I randomly sampled 30 melodies from a corpus of melodies (N = 481) from the Fifth Edition of the Berkowitz "A New Approach to Sight Singing" (**?**) in order to ensure a representative sampling of melodies that might be used in a pedagogical setting. After piloting the randomly sampled melodies on a colleague, I again randomly sampled half of this sub-set and then added in five more melodies that were not in the new set from earlier sections of the book in order to be more representative of materials students might find in the first two semesters of their aural skills pedagogy. I ran the survey from January 31st of 2019 until March 7th, 2019. The survey comprised of two sets of questions.

Six questions asked about the teaching background of respondents and can be found in TABLE 1. These questions were followed by asking participants to make five ratings over the 20 different melodies. The five questions can be found in TABLE 2. To encourage participation, two $30 cash prize was offered to two participants. The survey had questions that specifically were designed to gauge their appropriateness for use in a melodic dictation context. Participants were recruited exclusively online and all provided consent to partaking in the data collection as approved by the Louisiana State University Institutional Review Board.

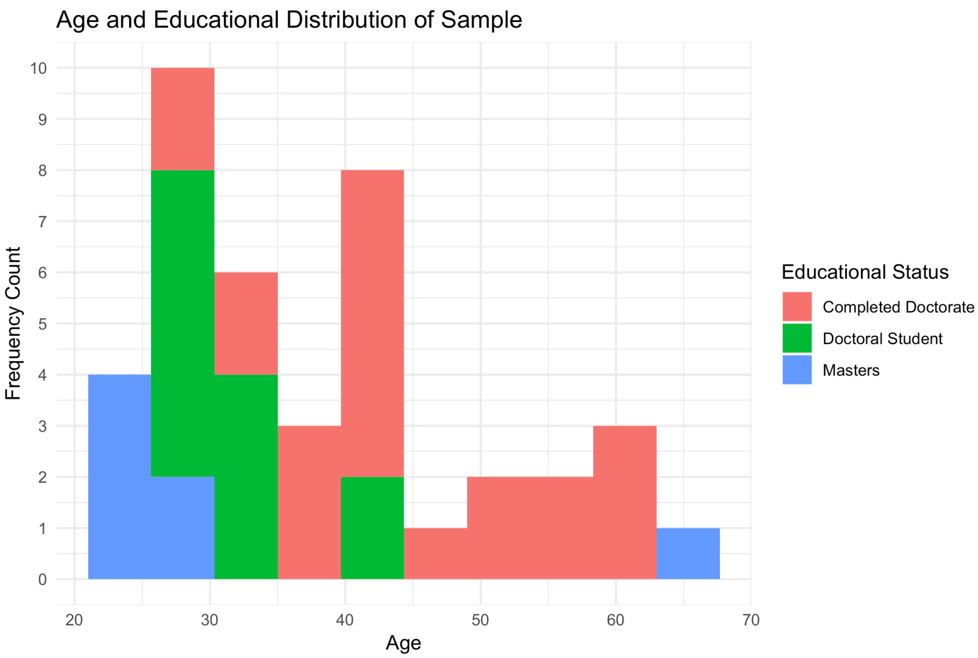


Figure 4.1: Demographic Breakdown of Sample

The table below contains the questions used in the demographic questionnaire.

| |
|---|
| 1. What is your age, in years? |
| 2. What is your educational status? (e.g. Master's Student, Doctoral Student, Completed Doctorate) |
| 3.How many years have you been teaching Aural Skills at the University level? Please do not include any Music Theory clas |
| 4. Which type of syllable system do you prefer to use? (e.g. Movable-Do, Fixed-Do, La-Based Minor, Numbers) |
| 5. On which instrument have you gained the most amount of professional training? (e.g. Piano, Voice, Marimba, Flute) |
| 6. What is the title of the last degree you received? (e.g. DMA Piano Pedagogy, PhD Music Theory, BA Music) |
| 7. At what institution are you currently teaching? If you are not currently teaching, but have taught in the past, please list |

The table below contains the questions regarding the ratings of the melodies. Participants either responed using ordinal categories or moved a slider that sat atop a 100 point scale.

| |
|---|
| 1. During which semester of Aural Skills would you think it is appropriate to give this melody as a melodic dictation? |
| 2. How many times do you think this melody should be played in a melodic dictation considering the difficulty you noted in |
| 3. Please rate how difficult you believe this melody to be for the average second-year undergraduate student at your institu |
| 4. Please rate this melody's adherence to the melodic grammar of the Common Practice Period. The far left should indicat |
| 5. Is this melody familiar to you? |

Of the respondents, the average amount of years teaching aural skills was 8.76 years ($SD = 7.60, R : 1 - 29$). I plot the breakdown of the respondant's age, educational status below in 4.1. Of the 40 repondants, all reported used some sort of moveable system other than 2 who used a fixed system. The sample represented over 30 different institutions.

Overall, the sample seems to reflect a wide range of experience of teaching aural skills. The sample has both younger and older individuals, as well as a range of experience. In the figures below, I list the 20 melodies sampled.

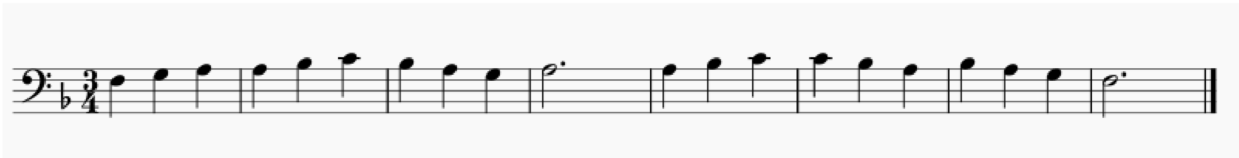

Figure 4.2: Berkowitz 3 | Rank 1

Figure 4.3: Berkowitz 9 | Rank 2

Figure 4.4: Berkowitz 26 | Rank 3


Figure 4.5: Berkowitz 59 | Rank 4


Figure 4.6: Berkowitz 70 | Rank 5

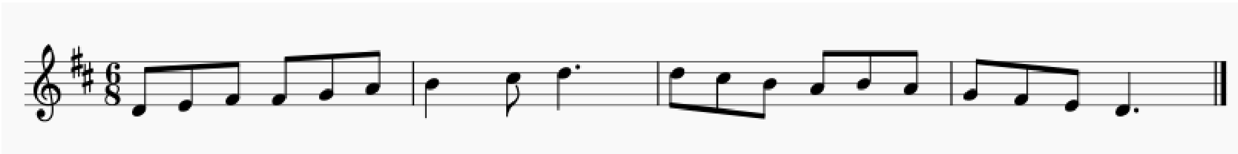

Figure 4.7: Berkowitz 74 | Rank 6

Figure 4.8: Berkowitz 75 | Rank 7

Figure 4.9: Berkowitz 88 | Rank 8

Figure 4.10: Berkowitz 156 | Rank 9

Figure 4.11: Berkowitz 282 | Rank 10

Figure 4.12: Berkowitz 294 | Rank 11

Figure 4.13: Berkowitz 312 | Rank 12

Figure 4.14: Berkowitz 334 | Rank 13


Figure 4.15: Berkowitz 379 | Rank 14

Figure 4.16: Berkowitz 382 | Rank 15


Figure 4.17: Berkowitz 417 | Rank 16

Figure 4.18: Berkowitz 607 | Rank 17

Figure 4.19: Berkowitz 622 | Rank 18

Figure 4.20: Berkowitz 627 | Rank 19


Figure 4.21: Berkowitz 629 | Rank 20

### 4.2.2 Agreement Among Peagogues

In order to assess the degree to which pedagogues agree on a melody for melodic dictation, I first plot the mean ratings for each melody across the entire sample along with their standard error of the means in Figure 4.22. The x axis uses the rank of the melodies, not their index position in the Berkowitz textbook. I chose to use use this rank order metric as the number of a melody in a textbook is presumed to be best conceptualized as an ordinal variable. For example, it would be correct to assume that Melody 200 is more difficult than melody 2, but not by a factor of 200.

From Figure 4.22, there is a clear, increasing linear trend from ratings of melodies being less difficult to more difficulty across the sample. Using an intraclass coefficient calculation of agreement using a two-way model (both melodies and raters treated as random effects), the sample reflects an interclass correlation coefficeint of .799. According to **?**, this reflects a good degree of agreement between raters. This trend across the sample appears in the opposite direction when plotting the mean values to the fourth question in Figure 4.23 from the survey reflecting the melody's adherence to the melodic grammar of the Common Practice period.

While similar trends appear here, yet in the opposite direction as expected, there is a clear breaking of linear trend in the far right potion of the graph that shows melodies that were sampled from the chapter of the Berkowitz that contains atonal melodies. Using an intraclass coefficient calculation of agreement using a two way model both melodies and raters treated as random effects, the sample reflects an interclass coeffiecent of .65, which according to **?** indicates a moderate degree of agreement among raters.

Both the trends from Figure 4.22 and Figure 4.23 occur in the opposite direction. As the index or rank of the melody increases, so does the difficulty for the rating as would be expected. As the index or rank of the melody increases, its adherence to subjective ratings of melodic grammar of the Common Practice period also decreases. Taken together, I ran a correlation on every one of the twenty melodies between a single rater's judged difficulty and its judged adherence to tonal expectations of the common practice era.[1] The correlations for all 20 melodies are plotted here in Figure 4.24. From this chart, we see higher degrees of correlation between difficulty and tonality.

Overall, the sample exhibited an acceptable degree of inter-rater reliability as measured by the interclass correlation coefficient. Plotting the respondent's answers across the textbook the melodies were taken from, with the book progressing from less to more difficult, it does appear that aural skills pedagogues tend to agree on how difficult a melody to be used in a dictation setting.

Central to my argument that, there appears to a linear trend of difficulty across the sample based on the melodies rank in the sample. In fact, although I presented the data above as ordinal, when I ran a mixed-effects linear regression predicting melody difficulty with both rank order as a variable as well as the melody index from the Berkowitz, the index model significantly outerperforms the rank order model.

[1] I chose not to pool ratings as that would violate the assumption of independence for correlation.

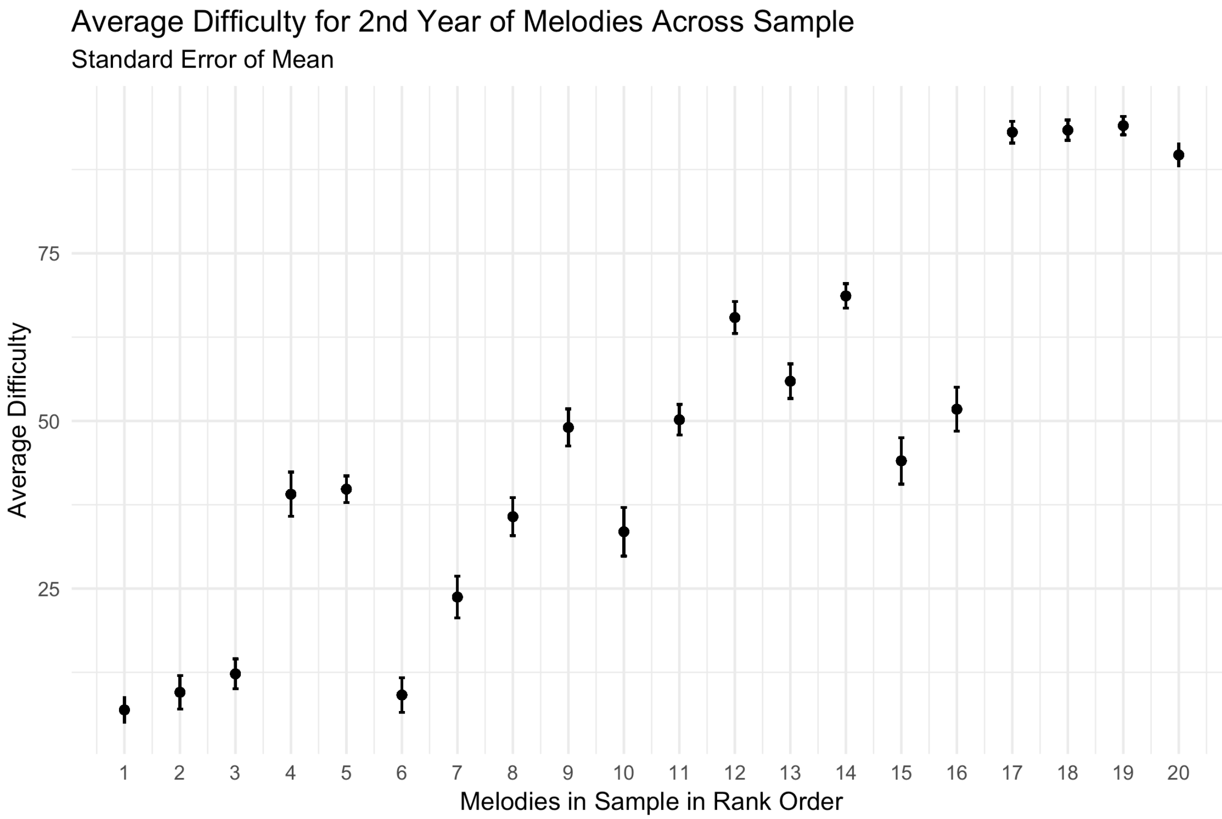


Figure 4.22: Average Difficulty

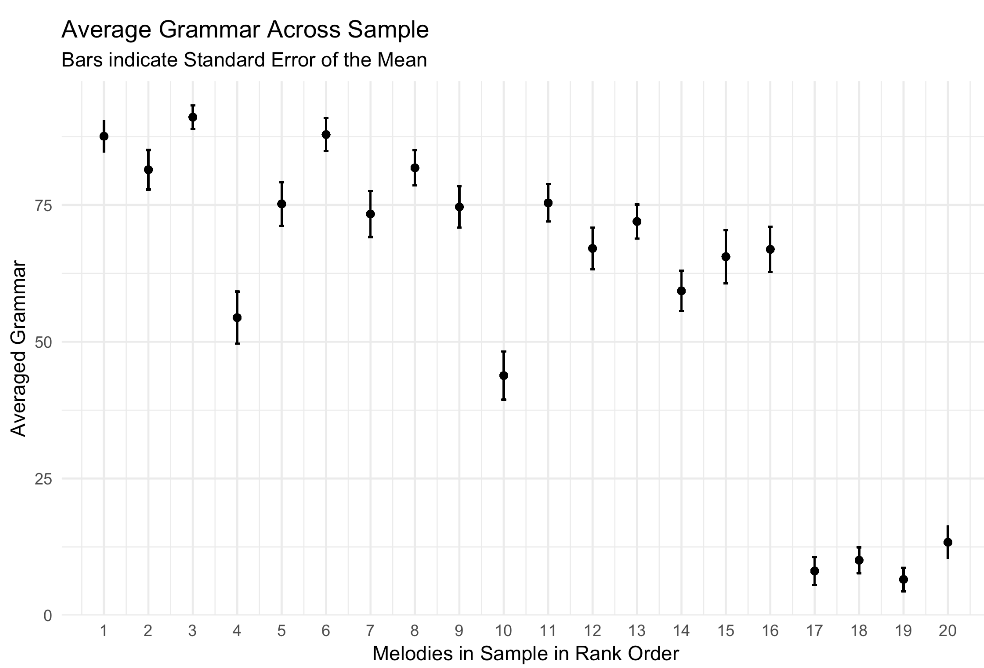


Figure 4.23: Average Grammar Ratings

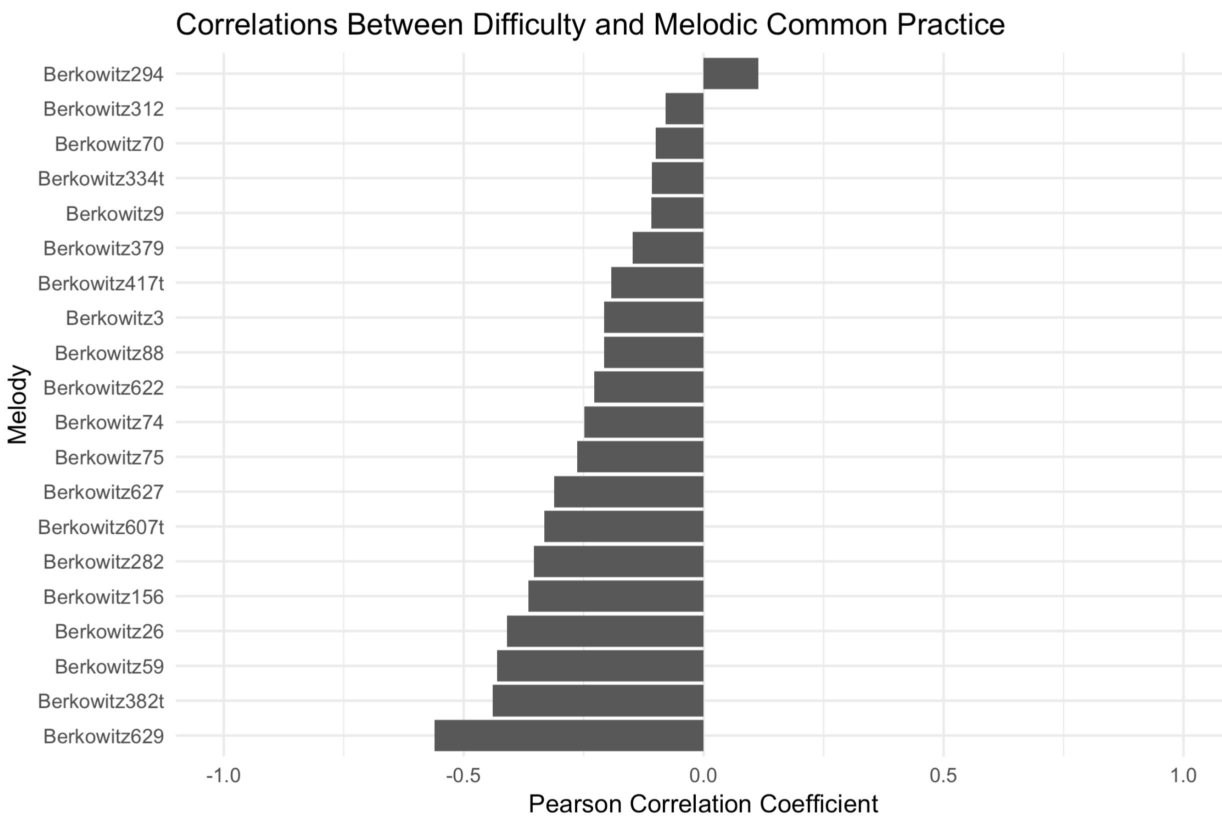


Figure 4.24: Strength of Relationship Between Difficulty and Subjective Tonal Grammar

- ADD STATS HERE

Taken together, both anecdotal and empirical evidence for this survey suggest that aural skills pedagogues tend to agree on how difficult a melody is for use in an aural skills setting. This sense of difficulty or complexity tracks as the book progresses, but to attribute the cause of a melody being difficult as its position in the book would be putting the cart before the horse. Having now formally established this almost intuitive notion, the remaining portion of this chapter investigates how computationally derived tools can be used to model these commonly held intuitions. In order to provide a sense of validity to the measure, I carry forward ratings from the survey reported and use the expert answers as the ground truth for the the resulting models.

## 4.3 Modeling Complexity

The ability to quantify what theorists generally agree to be melodic complexity depends on distilling complexity into its component parts. Earlier, when comparing melodies X and Y, some of the features put forward that might contribute to complexity were features such as note density, the melody's rhythm, what scale the melody draws its notes from, and how tonal the melody might be perceived. Some combination of these component features presumably make up the construct of complexity.

Attempting to use features of a melody to to predict how well a melody is remembered has a long history. In 1933, Ortmann put forward a set of melodic determinants that he asserted predicted how well a melody was remembered. These features such as a melody's repetition, pitch-direction, contour (conjunct-disjunct motion), degree, order, and implied harmony (chord structure) were deemed to affect the melody's ability to be remembered (**?**).

Since Ortmann, pedagogues such as Taylor and Pembrook have expanded on this research, finding significant effects of musical features such as length, tonality, as well as type of motion as well as an effect of experimental condition (**?**). Following up on Taylor's investigation, **?** found evidence corroborating Ortmann's initial claims that his four major determinants (repetition, note direction, conjunct-disjunct motion, degree of disjunctivness) had a significant main effects on an individual's ability to take dictation, yet note that these values do not exhaustively explain the findings. In their discussion they also note the problems of completely isolating the effects certain musical features as when you change one parameter, others are also subject to change. When looking at changes in structural elements of melodies, there is a collinearity issue among features. Not only does this problem exist within features of melodies, but also among participants. In reflecting on other factors that might contribute to their results, the authors note

> Clearly, a complete hierarchy of determinants would constitute a very long list, because not only would the many melodic structures be included, but also their interactions with subject and environmental variables. The ones included in the present study (musical experience, melodic carryover, and response method) provided evidence that the melodic determinants are not constant; rather, they vary as a function of the subject and environmental factors, which in turn can have significant effects on music discrimination and memory. (p. 33)

The authors later in the article go on to stress that future work should both replicate their findings as well as expand their modeling parameters. They call for both a larger sample, a broader spectrum of musical experiences, and to investigate more musical features.

Since then, some, but not many researchers have employed using features of the melodies to predict a behavioral measure in experimental settings. Not using as extensive of a battery as Ortman, Taylor, or Pembrook, researchers in music psychology such as as **?**, **?**, **?**, **?** have used the number of notes in a melody as a successful predictor of difficulty in melodic perception and discrimination tasks. Expanding on just using frequency of note counts, **?** instead of looking at single measures of melodic complexity, addressed the melodic collinearity issue noted by Taylor and Pembrook by using data reductive techniques to derive a single complexity measure found to be predictive in their statistical modeling deriving these measures from the FANASTIC toolbox (**?**). Following this research, **?** also incorporated a similar measure of complexity in their model of leitmotiv recognition in which predicted recall rates.

Each of these examples operationalizes some feature of the melody with a quantitative, numerical proxy that is assumed to be able to be mapped to perception. Ortman referred to them as determinants, others such as Müllensiefen refer to them as features (**?**). Since the word feature refers to a 'distinctive attribute', I will use this terminology throughout the rest of the chapter, though note that other terms have been used.

### 4.3.1 What Are Features?

A feature can be either a quantitative or qualitative observable feature of a melody that is assumed to be perceptually salient to the listener. Features are often difficult to quantify with the traditional tools of music analysis. Often, these features come inspired from other domains like computational linguistics.

To given an example of a feature that is not related to just the number of notes, perhaps one of the most popular features in perception research in recent decades is the normalized pairwise variability index or nPVI. The nPVI began as a measure of rhythmic variability in language (**?**). Shown below, the nPVI quantifies the amount of durational variability in language. It works by comparing the variability of vowel length compared to syllable length

$$nPVI = 100 * [\sum_{k=1}^{m-1} |\frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2}/(m-1)]$$

where $M$ is the number of vowels in an utterance and $d_k$ is th duration of the $k^{th}$ item. (**?**)

In linguistics, the nPVI has been used to delineate quantitative differences between stress and syllable timed languages. Recently in the past decade, music science researchers have used the nPVI to attempt to investigate claims about the relationship between speech and language (**???**). While results are mixed regarding the nPVI's predictive ability and there have been recent calls to limit the measure's use (**?**), it does serve as a very good example of a computational derived measure. Just like summarizing the range of a melody by subtracting the distance between the lowest and highest notes, the nPVI summarizes a phrase and importantly assumes that this measure is representative of the entire phrase the calculation was performed upon.

In computational musicology, features of melodies can generally be classified into two main types: static and dynamic features. Static features compute a summary measure over the entire melody while dynamic features calculate values for each event onset in a melody. One of the most complete set of static computational measures as applied to music perception come from Daniel Müllensiefen's' Feature ANalysis Technology Accessing STatistics (In a Corpus) or FANTASTIC toolbox (**?**). According to FANTASTIC's technical report,

> "FANTASTIC is a program…that analyzes melodies by computing features. The aim is to characterise a melody or a melodic phrase by a set of numerical or categorical values reflecting different aspects of musical structure. This feature representation of melodies can then be applied in Music Information Retrieval algorithms or computational models of melody cognition." (pp. 4)

Drawing from fields both central and peripheral to music science, FANTASTIC computes a collection of 38 features to analyze features of melodies and joined a large and continuing tradition of analyzing music computationally (**?**, ; **?**; **?**; **?**; **?**; **?**; **?**). Additionally, FANTASATIC also provides a framework for comparing the features of a melody with a parent corpus from which the melody is assumed to belong similar to a sample-population relationship.

### 4.3.2 Back to the Classroom

Returning to the Aural Skills classroom, many of these features can be used to approximate the previously established intuitions of complexity as agreed upon by theorists. Below in Figure XXXX, I plot the the mean difficulty and grammar ratings given by experts for each melody in the experimental sample against


Figure 4.25: FANTASTIC and Expert Ratings

each the output of FANTASTIC's features by correlating the two measures. Additionally, XXX displays the five strongest posititve and negatively correlated features of FANTASTIC's output with the ground truth, expert ratings.

Feature

Difficulty

Grammar

i.abs.std

0.8858728

-0.8205238

i.abs.mean

0.8737730

-0.9077636

step.cont.loc.var

0.8680327

-0.7364870

i.entropy

0.8516330

-0.7359731

p.entropy

0.8397031

-0.7209406

d.median

-0.1919713

0.2224165

d.eq.trans

-0.2000724

0.0391843

mean.Yules.K

-0.4341525

0.3988128

tonalness

-0.4778019

0.4435525

mean.Simpsons.D

-0.5656707

0.5033565

From Figure 4.25 and Table HERE, there are some features that share a strong relationship with the ground truth of the expert intuitions. The top five features that correlate most strongs with the expert ground truths are related to the intervallic content of a melody. The first two features, `i.abs.std` and `i.abs.mean` are derrived measures using absolute interval distance computations. The other top three features, `step.cont.loc.var`, `i.entropy`, and `p.entroy` are related to entropy measures. Of the negatively correlated features, two linguistically derrived measures `mean.Yules.K` and `mean.Simpsons.D` both correlate with percieved difficutly, as does a measure of `tonalnesss` which in FANTASTIC is based on the Krumhansl key profiles (**?**).

One problem in tackling this problem is that although many of these variables correlate strongly with our target variables, both our grammar and difficulty ratings, one aspect not apparent in this analysis is the correlation between each of the features. In order to demonstrate this, in Figure 4.26 I visualize how all the ten features from HERE correlate with one another with mode additionally included to highlight the breakdown of the corpus.

Among these variables, we see that there is a very high degree of correlation between many of the variables. For example, the two features inspired from linguistics– `mean.Yules.K` and `mean.Simpsons.D` – exhbit an alarming degree of correlation. We also see in this dataset evidence of the inappropriateness of including some variables such as `d.median`, a measure relating rhythm.

Here in 4.26 we see computational evidence of claims made by **?** when reviewing exactly what features might contribute to the degree of difficulty from a melodic dictation. Given this collinearlity problem, it becomes very difficult to be able to isolate the effect of of one feature of the melody. One way to begin to understand these relationship would be to be to build statistical models that area able to partition covariance structures such as the general linear model when used in the context of multiple regression. Another method, as mentioned above, could instead take a more exhaustive, but less explanatory approach forward and follow past research (**??**), that uses data reductive techniques such as principal components analysis to obtain more accurate predictive measures of complexity.

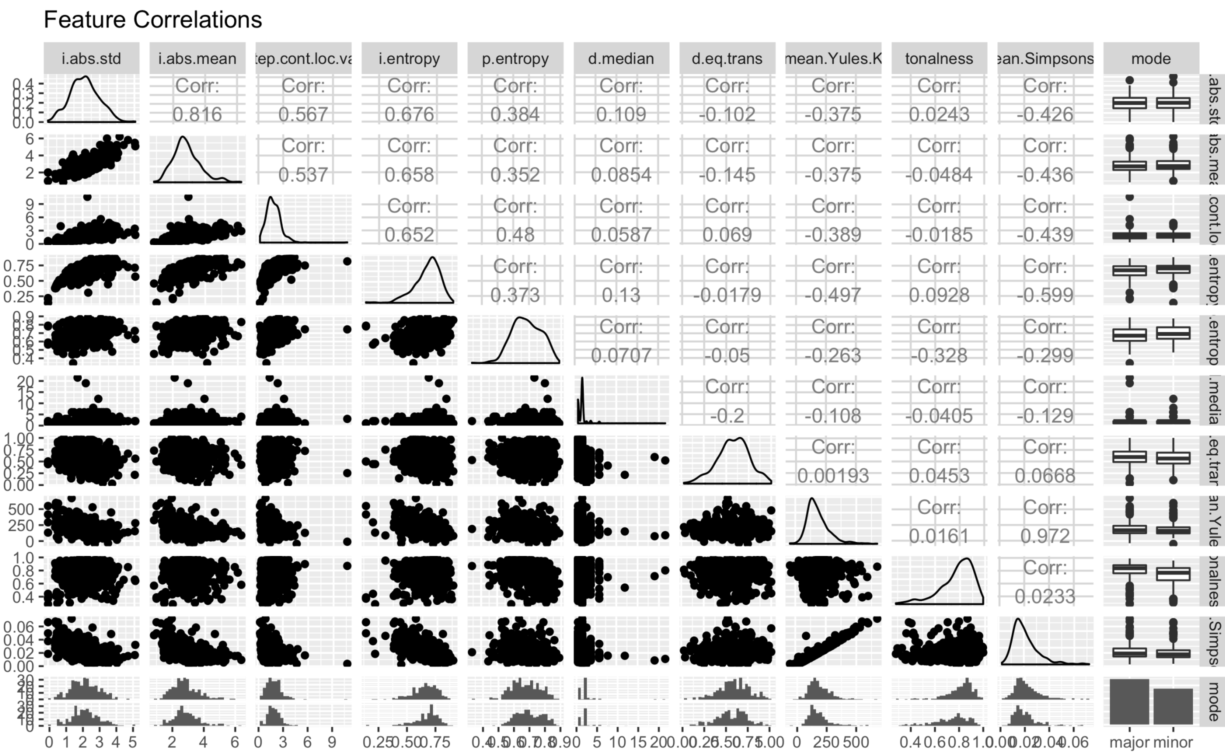


Figure 4.26: Problems of Melodic Collinearity

- Regression Analyses Here

Relating again back to its implication for aural skills pedagogy, the above analysis suggests that features as derrived from the FANTASTIC toolbox can provide a meaningful step forward in helping standardize the assesment of aural skills pedagogy. If pedagogues were able to employ tools such as the FANTASTIC toolbox, pedagogues could not only select melodies for their own work that is able to hold certain features constant, but use of this research could also be used generate melodies based on the desired dificulty parameter measures in order to design course curricula that would foster a more stable curricular path among students. Additionally, student could also work at slowly challenging themselves if this were to be incorporated into a pedagogical learning app or website.

Although this approach has been relatively successful at modeling expert ratings, using FANTASATIC's various linear combinations of these features does have important limitations. One of the most obviosu limitiations is that FANASATIC's measures tacitly assume listens experience melodies in some sort of perceptual suspended animation. Illuminating this problem using a more tangible example, again returing to melody's X and Y, when the full set of FANTASTIC features are computed on both, THIS FACT HERE ABOUT WHAT IS SAME. This computation arises from computing a summary measure over the melody and not modeling it in terms of real time perception. In order to have more phenomenologically appropriate model that incorporates computationally derived features, it is important to also consider dynamic models of music perception when modeling difficulty. Following up on another finding from this section, it also is worht of mention that the variables with the strongest predicitive powers– here– tend to be those associated with information content CHECK. In the next section, I explore how using a dynamic approach such as Marcus Pearce's Informaiton Dynamics of Music (IDyOM) model (**??**) might provide more insights into understanding the aural skills classroom.

### 4.3.3 Dynamic

The Information Dynamic of Music (IDyOM) model of Marcus Pearce is a computational model of auditory cognition (**?**). IDyOM is based on the assumption put forwar by Leonard Meyer that musical style can be understood as a complex network or probabilistic relationships that underly a musical style and implicitly understood by a musical community (**???**). Unlike measures from FANTASTIC, which calculate summary statistics based on melodic features, IDyOM works by calculating measures of expectancy of an event based on a predefined set of musical parameters that the model was was trained on. As mentioned in @ref(#intro), the IDyOM model relies on two important theoretical assumptions based on two neural mechanisms involved in musical enculturation: the statistical learning hypothesis and probabilistic prediction hypothesis. According to Pearce, the Statistical Learning Hypothesis (SLH) states that:

> musical enculturation is a process of implicit statistical learning in which listeners progressively acquire internal models of the statistical and structural regularities present in the musical styles to which they are exposed, over short (e.g., an individual piece of music) and long time scales (e.g., an entire lifetime of listening). p.2 (Pearce, 2018)

The logic here is that the more an individual is exposed to a musical style, the more they will implicitly understand its internal syntax and rules. The SLH leads the corroborating probabilistic prediction hypothesis which Pearce states as:

> while listening to new music, an enculturated listener applies models learned via the SLH to generate probabilistic predictions that enable them to organize and process their mental representations of the music and generate culturally appropriate responses. p.2 (Pearce, 2018).

Essentially IDyOM works by providing the model with a musical corpus that it assumes is representative of a genre, or musical style. This musical corpus then serves as training data to approximate either a listener or style's ground truth ASK MP?. After establishing this corpus, IDyOM then learns both long term and short term expectations of events using a variable-order Markov model[2] in order to best optimize its predictive abilities. The expectations that IDyOM calculates are based on a probability distribution of the proceeding events, which is then quantified in terms of informaiton content (CITE). As detailed in a summary review article on IDyOM by Pearce, IDyOM has been successful at predicting

> Western listeners' melodic pitch expectations in behavioral, physiological, and electroencephalography (EEG) studies using a range of experimental designs, including the probe-tone paradigm visually guided probe-tone paradigm a gambling paradigm, continuous expectedness ratings, and an implicit reaction-time task to judgments of timbral change.

Additionally, Peace notes some of IDyOM successes in modeling beyond expectation, including successes in modeling emotional experiences in music, recognition memory, perceptual similarity, phrase boundary perception and metrical inference. Importantly in reviewing IDyOM's capabilities regarding memory for musical pitches, Pearce also claims that

> A sequence with low IC is predictable and thus does not need to be encoded in full, since the predictable portion can be reconstructed with an appropriate predictive model; the sequence is compressible and can be stored efficiently. Conversely, an unpredictable sequence with high IC is less compressible and requires more memory for storage. Therefore, there are theoretical grounds for using IDyOM as a model of musical memory.

Peace notes four studies (**????**) that show that more complex melodies are more difficult to hold in memory. This theoretical assertion and select empirical findings have important ramifications for the aural skills classroom. In a dictation setting, melodies that are more expected should tax memory less, thus making them easier to remember and dictate. If I assume that more expected melodies are easier to remember, then it follows that the information content measures of expectedness can then be used as a stand in measure of melodic memory. This notion is not new to music psychology and was discussed by David Huron relating exposure to musical material as following similar laws to the the Hick-Hyman hypothesis

[2] 43-46

(??) which Huron paraphrases as "processing of familiar stimuli is faster than processing of unfamiliar stimuli" (?, pp. 63)" which now a decade later can be further investigate using tools from computational musicology. Combining the Hick-Hyman hypothesis together with the above statistical learning hypothesis and probabilistic prediction hypothesis, I then put forward a new hypothesis: the frequency facilitation hypothesis.

## 4.4 Frequency Facilitation Hypothesis

The frequency facilitation hypothesis (FFH) makes two important assumptions that rely on both the statistical learning hypothesis and the perceptual facilitation hypothesis. The first, as stated above, is that humans learn melodies via the means predicted by the statistical learning hypothesis. In line with Huron's reading of the Hick-Hyman Law, melodic information that people are more familiar with will consequently be more expected. More expected notes will tax memory less than unexpected notes. This assertion would also be predicted by the probabilistic prediction hypothesis. Thus, given a sequence any set of notes, the frequency facilitation hypothesis posits that the efficiency in which a melody is processed in memory is proportionally related to its degree of expectedness when quantified in information content. Specifically, measures of expectaiton derrived from computational models of auditory cognition like IDyOM should be able to serve as a proxy for melodic information.

This hypothesis generates testable predictions that can be investigated to verify its verisimilitude. Important to aural skills pedagogy, the primariy prediction from this hypothesis would be that melodic patterns that occur more frequently in a corpus will be be easier to remember than those occurring less frequently. These frequency patterns should then directly relate to the amount of information content caclulated by IDyOM. If this relationship does exist, then it can be used to create strategies that would then create a more linear path to success for students learning to take melodic dictation.

In the final section of this chapter, I investigate this claim by conducting an analysis on a corpus of sight singing melodies to demonstrate this claim. I then take the findings from this corpus analysis and how how it can be applied in the aural skills classroom.

### 4.4.1 Corpus Analysis

In order to investigate frequency facilitation hypothesis, I conducted a corpus study using N = 622 melodies from the above used Fifth Edition of the Berkowitz "A New Approach to Sight Singing" (?). The FFH predicts that more frequently occuring patterns will result in lower information content– a general (tautological) byproduct of quantifying muscial feature tokens and doing computations with IC– and that these lower informaiton content measures, when quantified, will be able to predict load on memory. In order to examine this I first extracted a series of the most frequenltly occuring melodic tri grams from the Berkowitz corpus after transposing each melody to C major via the `solfa` tool in humdrum. I plot the resuling distributions of the top 1000 patterns of each fixed order predictions below in 4.27. For each distribution, I also plot the top and bottom ten tri-grams from the tri-grams in table XXXX..

From 4.27, we see that when plotted in terms of their frequency distributions, a small amont of the patterns make up for a very large the distribution of the corpus. As evident from HERE, we see that with the addition of more tokens added to the m-grams and also is a visual representation of why and how statistical predicitons become more unreliable with higher order predictions. Intuitivly, melodic patterns from the high frequency distribution of the table would seemingly be easier to remember and then dictate than those from the tails of the distributions.

Following up on this analysis, I then trained an IDyOM model on the same corpus of melodies and thus was able to calculate the average information content the opening bi, tri, and quint grams melodies in this corpus. In this computation, I explicitly assume that the underlying corpus of data is representative of an indivual's perosnal expectations of musical mateiral.

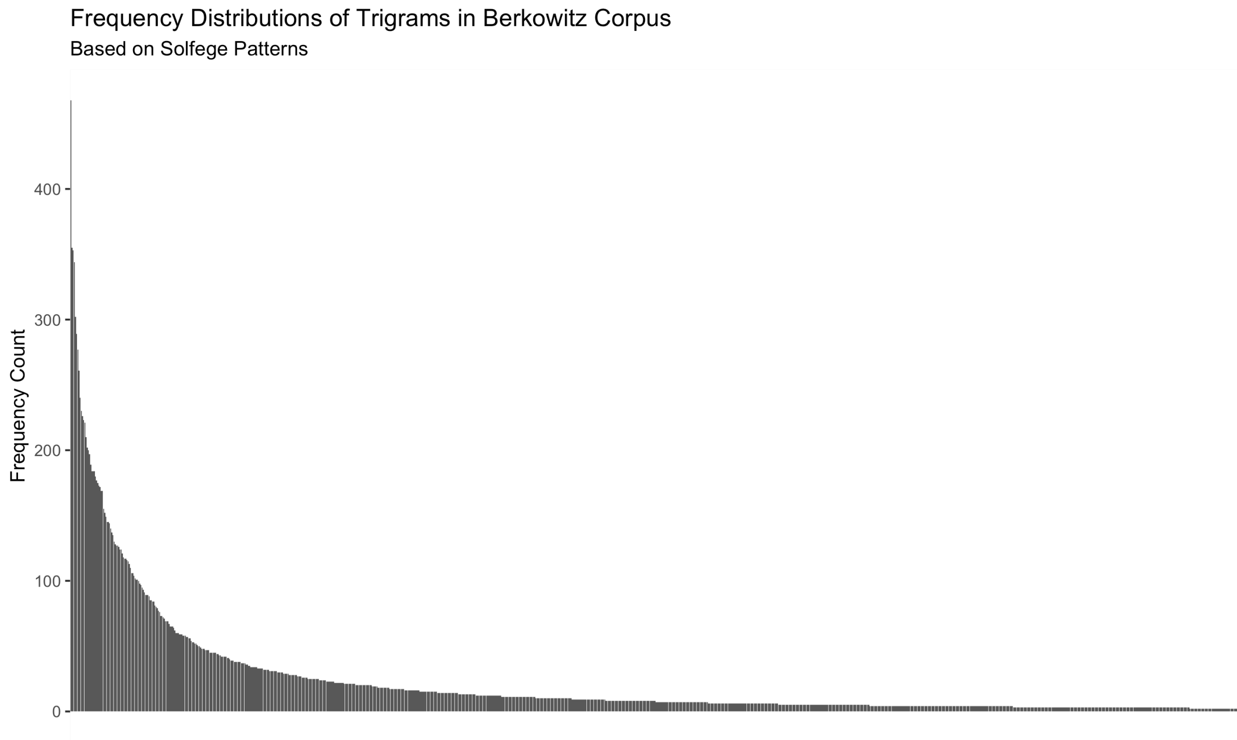


Figure 4.27: Distribution of m-grams

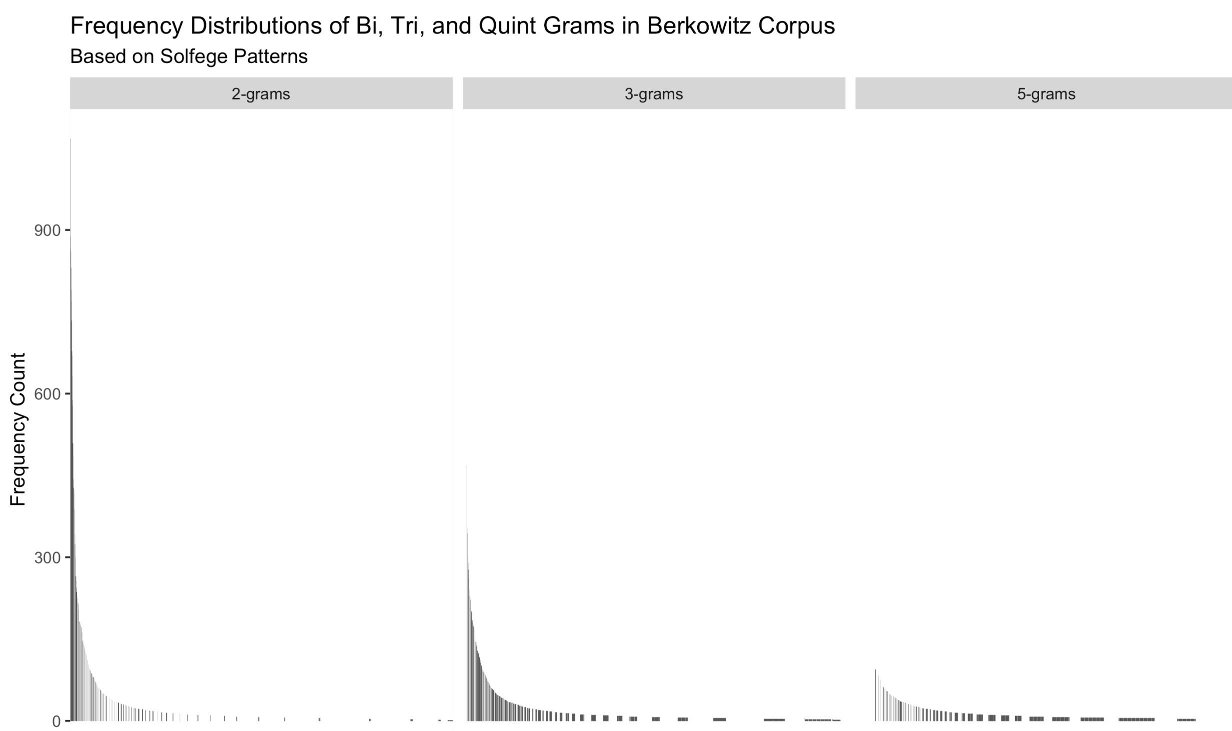


Figure 4.28: Less Predictive Power

From these computations, we can see that almost tautolitcally, that tri-grams with higher cumulative information content appear with lower frequency and m-grams with lower cumnulative informaiton content occur more frequently.

- HERE

This observation may seem tautological, as this relationship would result from how information content is calculated (more expeted patterns have less information), but the novel assertion here is connecting the cumulative informaiotn content to ease of memory load. For example, in Figure 4.29, we see that when split into three sections, random samples from the the corpus when partiioned into three secitons, even just the opening of the first five notes of each melody increase per group.[3]

To visualize what this might look like in a melodic dictation context, we could imagine randomly sampling melodies from even smaller sections (warrented by seeing the blip of data in the first quartile). If quantified using informaiton content measures, these five grams would then fill up the finite bin of memory than that were more unexpected, or had more informaiton content associated with them. I visualize this difference in XXXX where I plot similar length five grams filling up the window of memory at different rates based on their cumulative information content.

Lastly, to further investigate this claim of cumulative informaiton content, I calculated the average informaiotn cnotent for each melody baesd on several measures then used data from the survey above and plotted it against the measures of expert ratings for difficulty for the classrom and found various measures of information content to be very good predictors of difficulty ratings. I believe that htis gives creedence to thinking more about using comptuational measures in designing appropriate curricular measures XXXX.

- Talk about correlations
- Sig test these
- Discusssion

### 4.4.2 Implications

This implicaiton would have direct implications for the aural skills classroom, specifically for melodic dictation, as if measures of information content as modeled by IDyOM could be used as a more reliable proxy for load on memory, then a more linear path to learning patterns could result in better stratgies for learning to take dictation. For example, the first implication of this could be that information content, could be a more accurate proxy for the limits of musical memory as opposed to using older measures asserted by the literature that follow in the George Miller 7 +- 2 tradition which attempts to logically subsitiute items in memory for musical notes. While researchers liek X an Y have claimed this note-item limit to be X or Y, as disscussed in SECTION ON PROBLEMS WITH CHUNKING, using measures of infomraiton content could provide an ecologically acceptable work around to the problems of chunking within music. For example, if this measure proved to be useful in pedagogical applications, pedagoges would have a very powerful tool to create curricla that was designed in a much more linear path.

One of the major challenges in both teaching and learning aural skills beyond the identification of scale degrees is then identifying them in a more ecological, melodic context. Presenting snippets of melodies could be then be used as a very small intermediate step in teaching melodic dictation where students can then exhibit more frequent successes in the aural skills classroom in trying to dictate progressively difficulty snippets. If they learn the more frequent ones first, they will find them easier, but more important begin to recognize these patterns in longer exercises. Instead of picking melodies for practice one-by-one, pedagogues could instead give students a large compendium of small dictation exercises that were orderd to increase in their melodic information content over the course of instruction. In this type of applicaiton, students would not be learning to increase their melodic informmation capacity limit per se, but could provide a valueable means to give students multiple, smaller attempts to learn to take dictation, rather than being overwhelmed with longer melodies that are given to study on the premise of more ecological valididty. This could be done from the

[3] This same trend is also apparent using a general linear model across the entire corpus. I chose to model it with three groups for more effective communication.

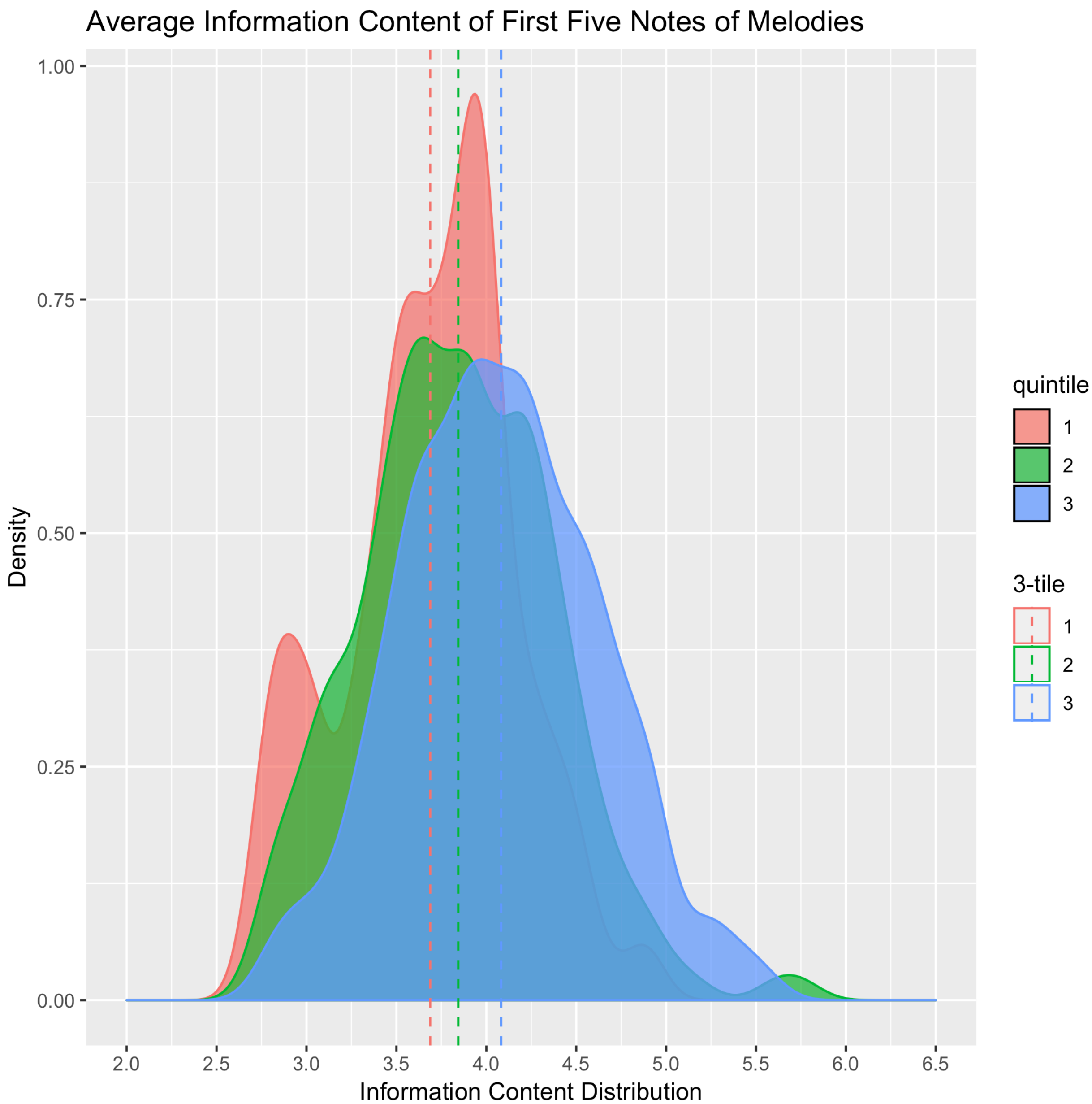


Figure 4.29: Average Information Content of First Five Notes of Melodies

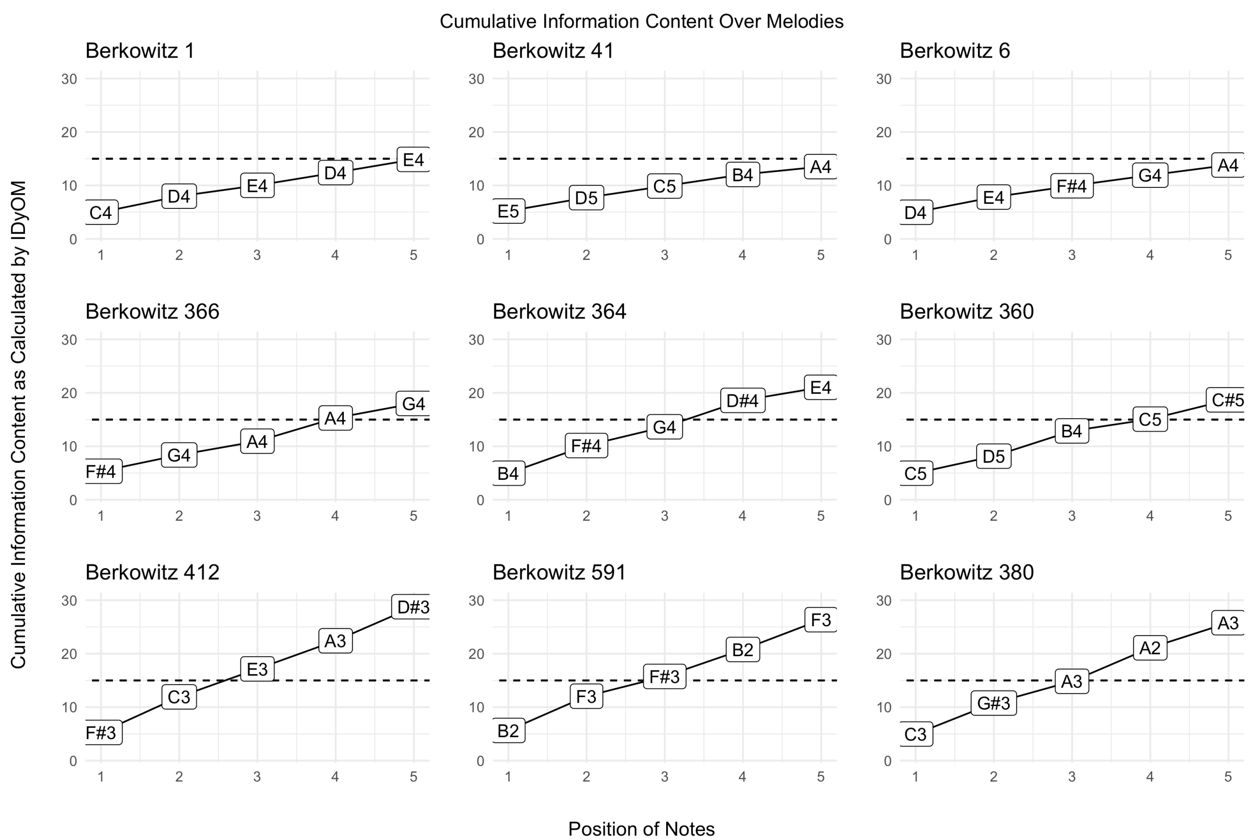


Figure 4.30: Cumlative Information Content in Melodic Incipts

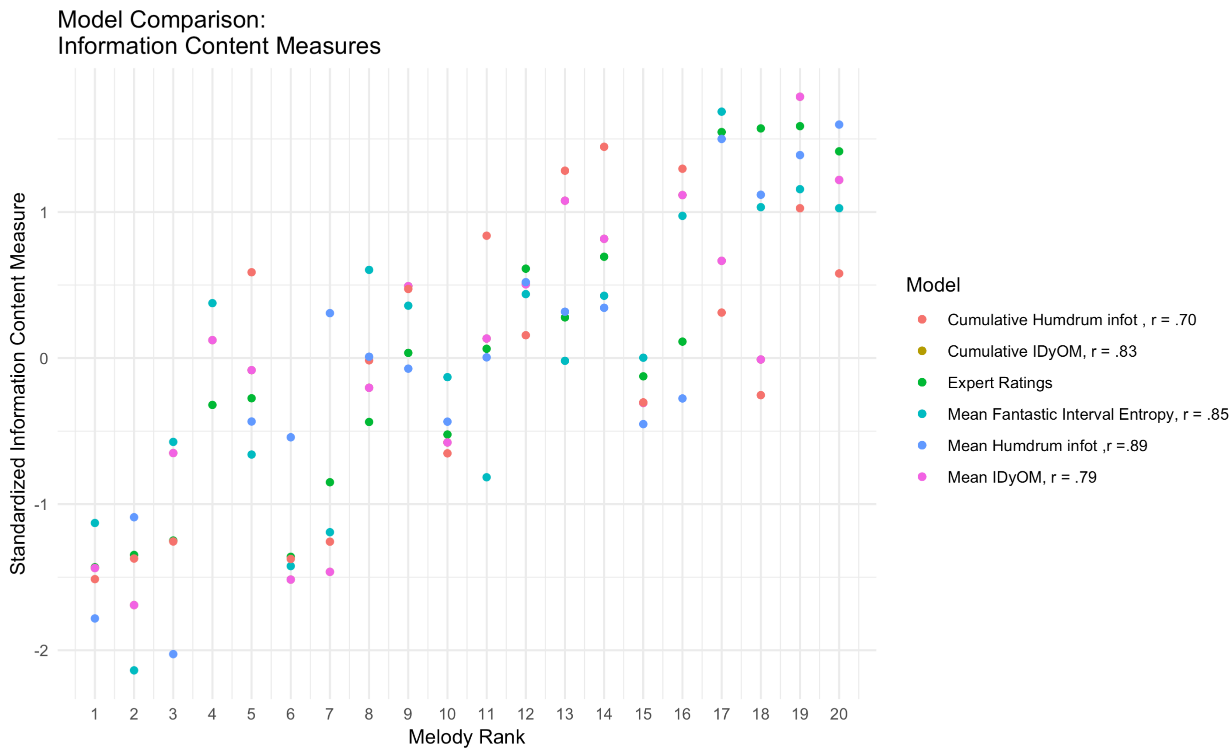


Figure 4.31: Model Comparison

level of scale degree identificaiton to that of full melodies. In order to demonstrate how this might work, in APPENDIX X I include a listing of 100 melodic incipits arraged in order of increasing information content. Future work should investigate this experimentally and look to model it using similar methodologies that have been employed in music psychology testing paradigms (IRT WAGER, WLF, PETER AND DANIEL ). Finally, if useful, this could and is type of modleing oculd also be used in future comptuational models of melodic dictation like THIS CHAPTER.

### 4.4.3 Limitations of FFH

This conceptualizaiton of calculating the information content of melodies is not without its limitations. One of the core assumptions to this approach is that statistical learning does in fact take place. While this assumption is ubiquitous in much the music psychology literature, statistical learning as a concept has been critiqued in other related fields and deserves mentioning. Statistical learning rests on the premise that organisms are able to implicitly learn and track the statistical regularities in their environments. In this case of auditory learning, there is research to assert this claim from both the field of implicit learning and statistical learning– two separate fields that have recently began to coalses CONWAY– as discussed by **?**. For example, extensive evidence as reviewed by FIND Cleeremans and Dienes (2008) show many examples of this, especially worht highlighting is that people have been show to learn variable order n-gram patterns (**?**).

Though this assertion is importantly contrasted by work such as **?**] who claim that explaining these phenomena as resulting in statistal leanring is not nessecary. Rather, Jamieson and XXX assert that employing memory models like that of MINERVA 2 (XXXX) can accuratly model behavioral patterns in individual responses without the theoretical framework of statistal learning. They instead note that similar results can result from individuals making similary jugments. This notation is important to highlight because as noted by **?**, statistical learning depend on the tacit assumption that people are performing real-time calculations on incoming stimuli in real time.

Another important caveat in the corpus analysis above is that the corpus analsis was done using fixed order search patterns[4], wheras the calculations from IDyOM are based on variable order Markov-Models. While differences in these computations might prove meaningful, only with future experimental evidence where we corrobrate with behvioral evidence would this be worht furhter looking into.

## 4.5 Conclusions

In this chapter I have demonstrated how tools from computational musicology can be used as an aide in aural skills pedagogy. After first establishing the extent to which aural skills pedagogues on various melody parameters, I then show how two families of computationally derived features can stand in for a pedagogues intuition. First, using the FANTASTIC toolbox, I show how different combinations of static abstracted features can help explain theorists agreed upon complexity. This first will help with selection of melodies and also provides insights as to which features of the melodies contribute most to perceived difficulty. Second, I demonstrated how assumptions derived from the IDyOM framework can serve as a basis for the intuitions of why smaller sequences of notes within melodies are more or less difficult to dictate. Using the logic that sequences that are easier to process are more expected, and that computed measures of information content can be used as a proxy for memory, I show that it follows that given the sequence of an N length melody, the ease of dictation that it loads on memory is relative to both its degree of quantified in terms of information content and link it back to the corpus by linking THAT to it's n-gram distributional frequency. This chain of thinking then allowed me to put forward a new sequence of melody segments that can be arranged, like other theory textbooks, in terms of their increasing complexity. I argue that using this smaller, snippet approach, will allow students to not be overwhelmed in their learning by taking a more linear path to dictation, before moving on to more more ecologically valid melodies. I finish by discussing how this might be implemented in the classroom.

[4] did a grep sort count on solfa

# Chapter 5

# Hello, Corpus

## 5.1 Rationale

One of the essential features of any scientific discovery is the ability to reproduce the finding. Given a new claim about reality, in order to be able to demonstrate that the claim is true, the new phenomea should reamin invariant when being repoduced under different conditions. If the phenomena satisfies pre-established criteria for causality can evidence be used to corrobrate theories that would predict this pheneomea (HUME, PEARL). This type of reasoning typically falls outside the realm of music theory, as studies involving music in the past few decades have generaly been associated with research in the humanities. One of the main objectives of research in the humanities– as opposed to that of the sciences– is to both challenge exisiting categories and subsequently create new framework of reality as we understand it (CITE). This division between the goals of those affiliated with studying music have not always been the case as noted by scholars like Allen Forte, who in his 1967 article notes that "In virtually any historic period one finds an interaction betweeen music and science and mathematics." (FORTE). This should come as suprise given that music, was one of the seven liberal arts during Roman times belonging to the quadrivium along with astronomy, geometry, and arithimitic (CITE). In fact, many disciplinary differences are more likely to result from geopolitical divides as how scholars concetuplize the study of music based on their location of study rather than the content and form of their research (PARNCUTT). Thus given th ebb and flow of relationship of music with so many other disciplines, incorporating epistomlogical and methodlogoical frameworks is nothing new in music research.

Returning to a phenomena's invariance under different conditions, one of the most effective ways to investigate claims about the state of reality is to reproduce previously made claims using new data. One of the most important contributions that a reseacher can make towards either bolstering or refuting claims about the nature of music and its resulting theories would be to generate more materials in which to examine previous claims under new conditions.

In order to accomplish this, in this chapter I introduce a new corpus of sight-singing melodies based on the pedagogical text "A New Approach to Sight Singing" (**?**). The corpus contains XXX monophonic melodies that have been digitially encoded in the kern format (**?**) and contain both melodies specifically composed for use in the Aural Skills classroom and examples of melodies from outside musical literature. After introducing the corpus, I compare the Sight-Singing corpus with the Essen Folk Song (**?**) collection in order to highlight variability between these musical corpora. I end by highlighting important considerations in the undelrying representaiotns of what the data represent and what these assumptions entail for future work in computational musicology.

## 5.2 History

The use of computers to study music has been been ongoing for over the past fifty years CITE. As catalouged by CITE, early approaches to using music to study computers begin in the mid 1960s and due to the high effort and cost of computation, projects pursued by researchers at this time tended to focus on questions that might have some sort of global relevance CITE. The use of computers to study music at this time was not by any means a sparse area of study as reflected in THIS BIBLIOGRAPHY by XXXX and throughout the second half of the 20th century, reserach in computational musicology grew in relation to the computing abilities afforded by the avalible technologies (CITE). During this time, not only was there progress made on computing power, many forms of developing new encoding frameworks were developed. As discussed by WIGGINS 1993, the design and development of these encoding frameworks impacts the degree that the systems can be asses based on two orthognal dimensions they identify in evaluating a framework: it's expressive completeness and structural generality. Considering what system is used to then encode musical information then becomes paramount given that the level of granularity that musical data in encoded will eventually determine the types of questions that could eventually be asked in an analysis. For example, data encoded in a MIDI or CHARM format is able to store mico-time variations in performance practice, which might lend it to being able to do performance based analyses on this data. This type of research has been done with CHOPIN PROJECT. If this data were instead to have been encoded in just using a frequency spectrum as would be stored in an MP3 or WAV file, this type of analysis could not be carried out as accuratly due to the task of automating the detection of pitch onsets.

On a higher level of abstraction, this problem of how a melody is encoded becomes exacerbated when considering meta-research issues such as the the tools-to-theories heuristic put forward by GIZO (CITE). GIZO makes the claim that much of both the novelty and authority given to the trajectory of a reserach path is determined by the tools a group decideds is valid, and not the generation of new data or theories. Contexutualizing this problem for digitial music encoding, again choosing how to represent the data reflects ontological and episotmological assupmtions about the data itself. Further, the technology used to be able ot query or test this data would provide an additional constraint on the analysis. Authors like XXX have noted this phenomena is fields like music informaiton retrevial, where much of the world has tended to focus on rhythmic issues such as beat tracking or onset dectection as opposed to VOCAL EXTRACTION due to the fact that it is easier to get this information from an audio signal. These issues are by no means settled, but as time progresses, moving between encoding formats becomes less difficulty with time. Some popular formats today include the Music Encoding Inititive, which according to their website is "a core set of rules for recording physical and intellectual characteristics of music notation documents expressed as an eXtensible Markup Language (XML) schema" (CITE), as well as MIDI (CITE), and kern (HUMDRUM). Not only is there variability in types of encoding, but additonally the tools availible to analyze digital music data vary as well. Popular analysis software consists of THIS PERSON's music21 writte in Python, David Huron's Humdrum (CITE), as well as technologies being developed by the SIMSSA project based in McGill under the direction of ICH. Despite differences the advantages between both types of encoding and tools use to analyze this data, parsers like that of KLAUS FRILERS are constanly being developed to serve as digital music's rosetta stone resulting in a current eco-system that allows for moving between options granted a compotent degree of computer programming experience.

While many of the encoding formats thoguhtou the past 50 years have fallen out of favor, the kern format of encoding data developed by David Huron has persisted as a choice for many computational musicologists since its initial development in 1994 (CITE). The kern format (often stylized as `**kern`) was developed in tandem with the Humdrum Toolbox for music analysis that according to Humdrum user guide is

> a set of command-line tools that facilitates musical analysis, as well as a generalized syntax for representing sequential streams of data. Because it's a set of command-line tools, it's program-language agnostic. Many have employed Humdrum tools in larger scripts that use PERL, Ruby, Python, Bash, LISP, and C++.

Humdrum files, unlike that of anything used in MEI are human readable and non-hierarchical, thus mirroring Western notated music's sequential time baed nature. Because of this, editing kern files using the humdrum toolset and humdrum extras developed by Craig Sapp (CITE) can be done with short, UNIX scripts as oppsed to similar analyses in music21. As humdrum is also sequential and text based, it is also able to encode formats typically not able to be supported by traditional formats such as THIS STUDY THIS STUDY. Since moving between digitially encoded ecoystems is not nearly as difficutl and much of encoding is can be left to the jusrdicition of the resarcher, I have chose to encode this datast using the kern format.

## 5.3 Berkowitz Corpus

In this next section I introduce a new corpus of melodies encoded in the kern format. The melodies come from the 5th edition of "a new approach to sight singing" written by Sol Berkowitz, Gabriel Fontrier, Leo Kraft, Perry Goldstein, and Edward Smaldone, CITE. In this verison, it includes XX melodies from the first and last chapter fo the book. The first Chapter contains melodies from five different sections and the fifth chapter comtaines "Melodies from the Literature" and is made up of four sections. Melodies from the first chapter have all been specifically composed for use in a sight singing context. Melodies from the fifth chapter are small excerpts from examples of both excerpts from Western Classical Music canon and traditional folk songs of various countries. The information for each melody is recorded in the meta-data of the kern file. In additon to having a key signature in each kern file, I have also added an explicit key to each kern file. Each section of the book contains melodies that would be considered tonal, except for melodies in the fifth section of the first chapter and intermittent melodies in the fourht section of the fifth chapter which contain atonal melodies. If a melody is decidedly atonal or written using a mode, this is document as well in the metadata. Atonal melodies are given the explicit key of C major so that they can be analyzed using tools from the humdrum extras toolbox and parsed as if they were part of a fixed do system. In totaly the corpus consists of XXXXXXX tokens. In Figure XXX see that there is even distirbution of keys. In comparison to other corpora, such as the canoncical Essen folk song datatbase, the Berkowtiz has less melodies, but melodies are typically LONGER, reflects more diversity of keys, and does not reflect music of a particular national culture. Contrasting the XXXXX tokens in the corpus, the Essen folk song ha XXXXXXX tokens.

- FIGURE KEY DISTRIBUTION

## 5.4 Comparison of Corpora

In order to givea brief overview of the corpus and contextualize it in the context of other corpora, in the next section I compare the Berkowtiz corpus to that of the Essen Folk song collection (CITE). I compare it to the Essen Folk Song Collection because imporantly close in being that they both were written for some sort of vocal performance. The Berkowitz corpus was specicificaly designed for pedagogical purposes whereas the Essen is more ecologically reflective of melodies originating from a diversity of sources. Though given that they are both generally vocal melodies and both come from Western sources, there presumably qwould be differences on between the two corpora on a large scale structure. A second, more important reason for comparing this corpus with the Essen is that the Essen is one of the more heavily cited corpora used in the field of comptuational musicology and often taken as a proxy to represent the underlying expectational structure of Western music. For example……. Much of this work makes claims about general level musical features, referenceing the Essen dataset as reference. In this context, the inderlying assumption in this inference is that the Essen is a sample of the larger population of Western music borrowing undelrying logic from frequentism as evident from the choice to examine these relationship like Huron did using frequestist statistics and the null hypothesis signifcance testing framework (Dienes chapter). While of course thi might be true,in order to have more evidence for this, would need to build more eviedence. This could come in the form of performing frequentist staistical tests on data to reject null looking at the probabily of the data given the hypothesis, tho on large datasets that is a bad idea. Or could be talkign about modeling these things in terms of Bayes, which would be modeling data lookign at probabily of the hypothesis given the data. Either way, providing more evidence for previous claims depends on, as noted above, finding new evidence for claims with new data.

Below I do a corpus analysis of both the Berkowitz an Essen on a descriptive level, then take a case study example and look into it where MELODIES DESCEND to see.

- Distribution of Solfa
- Distribution of Notes
- Distribution of bi-gram probabilities of them
- Then subtract the matricies to show where teh differnces in the bi-grams are

So knowing that they are similar in some aspects, want ot then ask if some of thes assumptions hold. For example, THEY argued that there is a descent in vocal tract. Found evidence using corpus study. THis was followed up and published by Shanahan. So if it is true, should see it again in this corpus.

In order to do that, ran THIS ANALYSIS over the humdrum data. And found that SENTENCE HERE. This is reflected in FIGURE HERE.

From this we can conclude that…. And this is an example of using new data to test invariance of claims under different conditions. But to return to question of assumptions, what population is are both of these corpora assumed to orginate from. And what, in fact are teh assumptions of other types of corpus studies. As said above, assumption here is that often in many studies corpora share an assumped sample population relationship. This is OK from descriptive point of view, but becomes difficult to then apply any sort of inferrential statisical tests. Requires answering a question of what are the bounds of hte population? In historical studies, this is never possible. No more Clara Schumann. In studies that caputured Essen over time, yes can be representiatve. But work in transmission of song reflects that just like culture evolves, so does music. This is discussed by SAaveage. So the more imporant question then becomes what does it mean to be able to replicate these findings and find what is invariant with another corpus. If this is not possilbe, epistomologically this prohibits from NHST. Frequentism is not the only way to conceptualize stiaticcal relationsip, could be Bayes. More often than not this also more reflects the intuition that peopel want of probabily of hypothesis given data rather than what NHST gives of P data given hypothesis. Concrete answer to this question is beyond the scope of this chapter, important to end by saying that people who do use corpus methodlogies need to be explicit about what they are assumign their corpora to represent.

# Chapter 6

# Experiments

Rationale

ITC

Have done all this and have not actually talked about dictation yet Clearly many factors contribte to this whole thing and need to be taken into a model Dictation is basically a within subjects design Experiment Get very ecological and dirty and run it

In this chapter I combine ideas put forward in chapter 2 and chapter 3 in an experiemnt looking at melodic dictaiton. I show how you can combine both tests of individual ability and musical features to predict score. Discuss new intricies of this kind of testing like scoring, relate back to classroom. And conclude with showing how if you cobmine aural skills, comptuational musicology, mixed effects, this is way forward.

Abstract

Despite its abundance in curricula in music conservatory settings, research on topics pertaining to aural skills is currently limited at best. While anthologies of materials for sight singing and dictation exist, the ways in which people learn melodies are not well understood. This problem is difficult to tackle given the amount of factors that may contribute to the process, such as the complexity of the melody, the degree of exposure needed to commit a melody to long-term memory, and individual differences in cognitive ability that have been shown to contribute to an individual's performance on musical tasks. Fortunately, literature exists in related areas that serve to inform which parameters might contribute to an individual's performance in a melodic dictation setting. This paper presents findings from an experiment (N=39) modeling performance on melodic dictation tasks using both individual and musical features. Results suggest tools from computational musicology as well as individual difference measures need further exploration in order to assess the degree to which various features contribute to melodic dictation performance and inform pedagogical practices.

## 6.1 Introduction

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

### 6.1.2 Dictation is basically a within subjects design Experiment

#### 6.1.2.1 Get very ecological and dirty and run it

Despite its near ubiquity in Conservatory and School of Music curricula, research surrounding topics concerning aural skills is not well understood. This is peculiar since almost any individual seeking to earn a degree in music usually must enrol in multiple aural skills classes which cover a wide array of topics from sight-singing melodies, to melodic and harmonic dictation– all of which are presumed to be fundamental to any musician's formal training. Skills acquired in these classes are meant to hone the musician's ear and enable them not only to think about music, but, to borrow Gary Karpinski's phrase, to "think in music" (Karpinski, 2000, p.4). The tacit assumption behind these tasks is that once one learns to think in music, these abilities should transfer to other aspects of the musician's playing in a deep and profound way. The skills that make up an individual's aural skills encompass many abilities, though are thought to be reflective of some sort of core skill. This is evident in early attempts to model performance in aural skills classes where C. S. Harrison, Asmus, and Serpe (1994) created a latent variable model to predict an individual's success in aural skills classes based on musical aptitude, musical experience, motivation, and academic ability. While their model was able to predict a large amount of variance (73%), modeling at this high, conceptual of a level does not provide any sort of specific insights into the mental processes that are required for completing aural skills related tasks. This trend can also be seen in more recent research that has explored the relationship between how well entrance exams at the university level are able to predict success later on in the degree program.

Wolf and Kopiez (2014) noted a multiple confounds in their study attempting to asses ability level in university musicians such as inflated grading, which led to ceiling effects, as well as a broad lack of consistency in how schools are assessing success within their students. But even if the results at the larger level were to be clearer, again this says nothing about the processes that contribute to tasks like melodic dictation. Rather than taking a bird's eye view of the subject, this chapter will primarily focus on factors that might contribute to an individual's ability dictate a melody. – note here on not causal, still descriptive to foreshadwo next chatpter

Melodic dictation is one of the central activities in an aural skills class. The activity normally consists of the instructor of the class playing a monophonic melody a limited number of times and the students must use both their ears, as well as their understanding of Western Music theory and notation, in order to transcribe the melody without any sort of external reference. No definitive method is taught across universities, but many schools of thought exist on the topic and a wealth of resources and materials have been suggested that might help students better complete these tasks (Berkowitz, Frontier, & Kraft, 1960; Cleland & Dobrea-Grindahl, 2013; Karpinski, 2007; Ottman, 1996). The lack of consistency could be attributed to the fact that there are so many processes at play during this process. Prior to listening, the student needs to have an understanding of Western music notation at least to the degree of understanding of the melody being played. This understanding needs to be readily accessible, since as new musical information is heard, it is the student's responsibility to, in that moment, encode the melody into either hold a chunk of the melody in short term memory or pattern match to long term memory so that they can identify what they are hearing and transcribe it moments later into Western notation. So no matter what, performing some sort of aural skills task requires both long term memory and knowledge for comprehension, as well as the ability to actively manipulate differing degrees of complex musical information in real time while concurrently writing it down.

Given this complexity of the task, as well as the difficulty in quantifying attributes of melodies, it is then not surprising that scant research exists on describing these tasks. Fortunately, a fair amount of research exists in related literature which can generate theories and hypotheses explaining how individuals dictate melodies. Beginning first with factors that are less malleable from person to person would be individual differences in cognitive ability. While dictating melodies is something that is learned, a growing body of literature suggests that other factors can explain unique amounts of variance in performance via differences in cognitive ability. For example, Meinz and Hambrick (2010) found that measures of working memory capacity (WMC) were able to explain variance in an individual's ability to sight read above and beyond that of sight reading experience and musical training. Colley, Keller, and Halpern (2017) recently suggested an individual's WMC also could help explain differences beyond musical training in tasks related to tasks of tapping along to expressive timing in music. These issues become more confounded when considering other recent work by Swaminathan, Schellenberg, and Khalil (2017) that suggests factors such as musical aptitude, when considered in the modeling process, can better explain individual differences in intelligence between musicians and nonmusicians implying that within the musical population. They claim there is a selection bias that "smarter" people tend to gravitate towards studying music, which may explain some of the differences in memory thought to be caused by music study (Talamini, Altoè, Carretti, & Grassi, 2017). Knowing that these cognitive factors can play a role warrants attention from future researchers on controlling for variables that might that might contribute to this process but are not directly intuitive and have not been considered in much of the past research. This is especially important given recent critique of models that purport to measure cognitive ability but are not grounded in an explanatory theoretical model (Kovacs & Conway, 2016).

Memory for Melodies

The ability to understand how individuals encode melodies is at the heart of much of the music perception literature. Largely stemming from the work of Bregman (1994), Deutsch and Feroe (1981), and Dowling (1978; 1971) work on memory for melodies has begun to lay the foundation for how people learn melodies. Initial work by Dowling suggested that both key and contour information play a central role in the perception and memory of novel melodies. Interestingly enough, memory for melodies tends to be much worse than memory for other stimuli such as pictures or faces noting that the average area under the ROC curve tends to be at about .7 in many of the studies they reviewed, with .5 meaning chance and 1 being a perfect performance (Halpern and Bartlett, 2010). Halpern and Bartlett also note that much of the literature on memory for melodies primarily used same difference experimental paradigms to investigate individual's melodic perception ability similar to the paradigm used in Halpern and Mullensiefen (2008).

Musical Factors

Not nearly as much is known about how an individual learns melodies, especially in dictation setting. The last, and possibly most obvious, variable that would contribute to an individual's ability to learn and dictate a melody would be the amount of exposure to the melody and the complexity of the melody itself. There is not much research on the first of these two points, other than an approximation of how many times the melody should be played in a dictation setting according to (Karpinski, 2007, p.100) that accounts for chunking as well as the idea that more exposure would lead to more complete encoding.

Recently tools have been developed in the field of computational musicology to help with operationalizing how complex melodies are. Both simple and more complex features have been used to model performance in behavioral tasks. For example Eerola, Himberg, Toiviainen, and Louhivuori (2006) found that note density, though not consciously aware to the participants, predicted judgments of human similarity between melodies not familiar to the participants.

- Harrison, Baker, Oura, Taylor and Pembrook, Ortman

Note density would be an ideal candidate to investigate as it is both easily measured and the amount of information that can be currently held in memory as measured by bits of information has a long history in cognitive psychology (Cowan, 2015; Miller, 1956) – Also Marcus. In terms of more complex features, much of the work largely stems from the work of Mullensiefen and his development of the FANTASTIC Toolbox (2009), a few papers have claimed to be able to predict various behavioral outcomes based on the structural characteristics of melodies. For example, Kopiez and Mullensiefen (2011) claimed to have been able to predict how well songs from The Beatles' album Revolver did on popularity charts based on structural characteristic of the melodies using a data driven approach. Expanding on an earlier study, Mullensiefen and Halpern (2014) found that the degree of distinctiveness of a melody when compared to its parent corpus could be used in order to predict how participants in an old/new memory paradigm were able to recognize melodies.

These abstracted features also have been used in various corpus studies (Frieler, Jakubowski, & Mullensiefen, 2015; Jakubowski, Finkel, Stewart, & Mullensiefen, 2017; Janssen, Burgoyne, & Honing, 2017; Rainsford, Palmer and Paine 2017) that again use a machine learning approach in order to explain which of the 38 features that FANTASTIC calculates can predict real-world behavior.

In addition to looking at individual features, or sets of features, as predictors, recent work by P. Harrison, Musil, and Mullensiefen (2016), Baker and Mullensiefen (2017) and the aforementioned Mullensiefen and Halpern (2014) study have used data reduction techniques, namely principal component analysis, to take measures that were successful in predicting behavioral outcomes and boil them down into a single measure of complexity that has had predictive power in modeling experimental performance. While helpful and somewhat explanatory, the problem with many of these approaches is that they take a post-hoc data driven approach with the assumption that listeners are even able to abstract and perceive these features. Doing this does not allow for any sort of controlled approach and without experimentally manipulating the parameters, which is then further confounded when using some sort of data reduction technique. – tho link here to other chapter This is understandable seeing as it is very difficult to manipulate certain qualities of a melody without disturbing other features. For example, if you wanted to decrease the "tonalness" of a melody by adding in a few more chromatic pitches, you inevitably will increase other measures of pitch and interval entropy. In order to truly understand if these features are driving changes in behaviour, each needs to be altered in some sort of controlled and systematic way while simultaneously considering differences in training and cognitive ability. – tho what would judea pearl say about the causal diagrams here????

- Write text here for second experiment if possible !!! (For MP )

AIMS

This paper presents findings from two experiments modeling performance on melodic dictation tasks using both individual and musical features. A pilot study was run (N=11) was used in order to assess musical confounds that might be present in modeling melodic dictation. Results of that pilot study are not reported here. Based on the results of this pilot data, a follow up experiment was conducted to better investigate the features in question.

The study sought to answer three main hypotheses:

1. Are all experimental melodies used equally difficult to dictate?
2. To what extent do the musical features of Note Density and Tonalness play a role in difficulty of dictation?
3. Do individual factors at the cognitive level play a role in the melodic dictation process above and beyond musical factors?

## 6.2 Methods

### 6.2.1 Participants

Participants Forty-three students enrolled at Louisiana State University School of Music completed the study. The inclusion criteria in the analysis included reporting no hearing loss, not actively taking medication that would alter cognitive performance, and individuals whose performance on any task performed greater than three standard deviations from the mean score of that task. Using these criteria two participants were dropped for not completing the entire experiment. Thus, 41 participants met the criteria for inclusion. The eligible participants were between the ages of 17 and 26 (M = 19.81, SD = 1.93; 15 women). Participants volunteered, received course credit, or were paid $10.

### 6.2.2 Materials

Four melodies for the dictation were selected from a corpus of N=115 melodies derived from the A New Approach to Sight Singing aural skills textbook by Berkowitz et. al (2005). Melodies were chosen based on their musical features as extracted via the FANTASTIC Toolbox (Mullensiefen, 2009). After abstracting the full set of features of the melodies, possible melodies were first narrowed down by limiting the corpus to melodies lasting between 9 and 12 seconds and then indexed to select four melodies were chosen that as part of a 2x2 repeated measures design including a high and low tonalness and note density condition. Melodies, as well as a table of their abstracted features can be seen in Table 1 and Figures 1—4. Melodies and other sounds used were encoded using MuseScore 2 using the standard piano timbre and all set to a tempo of quarter = 120 beats per minute and adjusted accordingly based on time signature to ensure they all sounded the same absolute time duration. The experiment was then coded in jsPsych (de Leeuw, 2015) and accesed through a browser offline with high quality headphones.

paney 2016 is timing

Table 1

Figure 1-4

### 6.2.3 Procedure

Upon arriving at the lab, participants sat down in a lab at their own personal computer. Multiple individuals were tested simultaneously although individually. Each participant was given a test packet which contained all information needed for the experiment. After obtaining written consent participants navigated through a series of instructions explaining the nature of the experiment and given an opportunity to adjust the volume to a comfortable level. The first portion of the experiment that participants completed was the melodic dictation. In order to alleviate any anxiety in performance, participants were explicitly told that "unlike dictations performed in class, they were not expected to get perfect scores on their dictations". Each melody was played five times with 20 seconds between hearings and 120 seconds after the last hearing. (CITE SIMILAR) After the dictation portion of the experiment, participants completed a small survey on their Aural Skills background, as well as the Bucknell Auditory Imagery Scale C (Halpern, 2015). After completing the Aural Skills portion of the experiment participants completed one block of two different tests of working memory capacity (Unsworth et al., 2005) and Raven's Advanced Progressive Matrices and a Number Series task as two tests of general fluid intelligence (Gf) (Raven et al., 1998; Thurstone, 1938) resulting in four total scores. – Cite here that talks about how one round of each is OK? After completing the cognitive battery, participants finished the experiment by compiling the self-report version of the Goldsmiths Musical Sophistication Index (Mullensifen et. al, 2014), the Short Test of Musical Preferences (Rentfrow & Gosling, 2003), as well as questions pertaining to the participants SES, and any other information we needed to control for (Hearing Loss, Medication). Exact materials for the experiment can be found here.

### 6.2.4 Scoring Melodies

Scoring Melodies were scored by counting the amount of notes in the melody and multiplying that number by two. Half the points were attributed to rhythmic accuracy and the other half to pitch accuracy. Points were not deducted for notating the melody in the incorrect octave. Points for pitch could only be given if the participant correctly notated the rhythm. For example, in melody 34 there were 40 points possible (20 notes * 2). If a participant were to have put a quarter note on the second beat of the third measure, and have everything else correct, they would have scored a 19/20. Only if the correct rhythms of the measures were accurate could pitch points be awarded. In cases where there were more serious errors, for example if the second half of the second bar was not notated, points would have been deducted in both the pitch and rhythm sub-scores. Both the first and second author scored all melodies independently and then cross referenced for interrater reliability. – change wording Using a single score intraclass correlation coefficient calculation = .96 which suggests a high degree of inter-rater reliability (McHugh, 2012).

— discussion on scoring in discussion

## 6.3 Results

### 6.3.1 Data Screening

Before conducting any analyses data was screened for quality. List wise deletion was used to remove any participants that did not have all variables used in modeling. This process resulted in removing four participants: two did not complete any of the survey materials and two did not have any measures of working memory capacity due to computer error. After list-wise deletion, thirty-nine participants remained. Effects of Melodic Features In order to investigate H1, that melodies would differ in their degree of difficulty based on melodic features, we ran a repeated measures ANOVA using the ez package in R (Lawrence, 2016). Relevant statistics from the model can be seen in Table 2.

- RM ANOVA TABLE

Subsequent models exploring possible exploratory covariance relationships using random slope models that used measures of working memory capacity, general fluid intelligence, and measures of musical training, none of which emerged as significant. Differences between melodies can be see below in Figure 5.

- FIGRE 5 BOX PLOT

## 6.4 Discussion

Here, we have investigated the extent to which both individual differences and abstracted musical features could be used to model results in melodic dictations. In order to examine H1, we ran a repeated measures ANOVA in order discern any differences in melody difficulty. As noted in Table 2, both a significant main effect of Tonalness and Note Density was found, as well as a small interaction between the two variables suggesting evidence supporting rejecting H2's null hypothesis. The interaction emerged from differences in melody means in the low density conditions with the melody with higher tonalness actually scoring higher in terms of number of errors.

— Add in all music education literature — Get Ortmann

While we expected to find an interaction, this condition (Melody 34) was hypothesized to be the easiest of the four conditions. With Melody 9 there was a clear floor effect, which was also to be expected as when we chose the melodies, we had no previous experimental data explicitly looking at melodic dictation to rely on. For future experiments, we will use abstracted features from Melody 9 as a baseline. The main effect of note density was expected and exhibited a large effect size. ( g = .46). While it would be tempting to attribute this finding exactly to the Note Density feature extracted by FANTASTIC, the high and low density conditions could also be operationalized as having compound versus simple meter. Given the large effect of note density, we plan on taking more careful steps in the selection of our next melodies in order to control for any effects of meter and keep the effects limited to one meter if at all possible.

Somewhat surprisingly, the analysis incorporating the cognitive measures of covariance did not yield any significant results. While other researchers have noted the importance of baseline cognitive ability (Schellenberg & Weiss, 2013), the task specificity of doing melodic dictation as we designed the experiment might not be well suited to capture the variability needed for any effects. Hence, this paper would not be able to reject H3's null hypothesis. Considering that other researchers have founding constructs like working memory capacity and general fluid intelligence to be important factors of tasks of musical perception, a more refined design might be considered in the future to find any sort of effects.

Taken as a whole, these findings suggest that aural skills pedagogues should consider exploring the extent to which computationally extracted features can guide the difficulty expected of melodic dictation exercises.

Need section on the classroom

This paper demonstrates that abstracted musical features such as tonalness and note density can play a role in predicting how well students do in tasks of melodic dictation. While the experiment failed to yield any significant differences in cognitive ability predicting success at the task, our future research plans to continue incorporate measures that others have deemed important. We next plan to replicate this experiment's design with different melodies that usesimilar features.

- Redo with mixed effects everything
- Special parapgraph points on how DV changes will change modeling
- not that even after all of this still need a computaitonal model to put it all together

# Chapter 7

# Computational Model

## 7.1 Levels of Abstraction

In his 2007 article *Models of Music Similarity* (?), Geraint Wiggins distinguishes between *descriptive* and *explanatory* models in describing the modeling of human behavior. Descriptive models assert what will happen in response to an event. For example, as discussed in the previous chapter, as the note density of a melody increases and the tonalness of a melody decreases, a melody may become harder to dictate. While the increase in note density is assumed to drive the decrease in dictation scores, merely stating that there is an established relationship between one variable and the other says nothing about the inner workings of this process. An explanatory model on the other hand not only describes what will happen, but additionally notes why and how this process occurs. For example, much of the work musical expectation demonstrates that as an individual's exposure to a musical style increases, so does their ability to predict specific events within a given musical texture (?).

Not only does more exposure predict more accurate responses, but many of these models of musical expectation derive their underlying predictive power from the brain's ability to implicitly track statistical regularities in musical perception (??). The *how* derives from the tracking of statistical regularities in musical information and the *why* derives from evolutionary demands; Organisms that are able to make more accurate predictions about their environment are more likely to survive and pass on their genes (?).

Wiggins writes that although there can be both explanatory and descriptive theories, depending on the level of abstraction, a theory may be explanatory at one level, yet descriptive at another. Using the mind-brain dichotomy, he asserts that the example of a theory of musical expectation could be explanatory at the level of behavior as noted above, but says nothing about what is happening at the neural level. Both descriptive and explanatory theories are needed: descriptive theories are used to test explanatory theories and by stringing together different layers of abstraction, we can arrive at a better understanding of how the world works.

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

In this chapter I introduce an explanatory model of melodic dictation. The model is inspired by work from both computational musicology and cognitive psychology. From computational musicology I draw on the work of Marcus Pearce's IDyOM (?) and from cognitive psychology I draw from Nelson Cowan's Embedded Process model of working memory (??) to explain the perceptual components. In addition to quantifying each step, the model incorporates flexible parameters that could be adjusted in order to accommodate individual differences, while still relying on a domain general process. By relying on cognitive mechanisms based in statistical learning, rather than a rule based system for music analysis (**????**) this model allows for the heterogeneity of musical experience among a diversity of music listeners.

## 7.2 Model Overview

The model consists of three main modules, each with its own set of parameters:

1. Prior Knowledge
2. Selective Attention
3. Transcription and Re-entry

Inspired by Bayesian computational modeling, the *Prior Knowledge* module reflects the previous knowledge an individual brings to the melodic dictation. The *Selective Attention*– somewhat akin to Karpinski's extractive listening– segments incoming musical information by using the window of attention as conceptualized as the limits of working memory capacity as a sensory bottleneck to constrict the size of musical chunk that an individual could to transcribe. Once musical material is in the focus of attention, the *Transcription* function pattern matches against the *Prior Knowledge's* corpus of information in order to find a match of explicitly known musical information. The *Transcription* function will recursively truncate what musical information is in *Selective Attention* if no match is found. In addition to *Transcription*, there is also a *Re-entry* function that will restart the entire loop. This process reflects, but does not actually mirror the exact cognitive process used in melodic dictation, yet seems to be phenomenologically similar to the decision making process used when attempting notate novel melodies. Based on both the prior knowledge and individual differences of the individual, the model will scale in ability, with the general retrieval mechanisms in place. The exact details of the assumptions, parameters, and complete formula of the model are discussed below.

## 7.3 Verbal Model

Below I describe my model's assumptions, parameters, as well as the steps taken when the model is run. After detailing the inner workings of each of the assumptions and the modules, described in roughly the order that they occur, I present the model using psudeocode with the terminology described below. I discuss the issues of assumptions and representations as they arise in describing the model.

### 7.3.1 Model Representational Assumptions

In order to write a computer program that mirrors the melodic dictation process, how the mind perceives and represents about musical information must be defined *a priori*. Before delving into questions of representation, this model assumes that the musical surface[1] as represented by the notes via Western musical notation are salient and can be perceived as distinct perceptual phenomena. Although there is work that suggests that different cultures and levels of experience might not categorize melodic information universally (**?**), other work suggests that experiencing pitches as discrete, categorical phenomena is categorized as a statistical human universal (**?**). For the purposes of this model I assume that individuals do in fact perceive the musical surface similarly to the written score.

Knowing that it is melodic information or melodic data that needs to be represented, the question then becomes what is the best way in which to represent it. This issue becomes increasingly complex when considering literature suggesting that the human mind represents musical information in a variety of different forms (**??**).

For the purposes of this model and further examples I choose to represent musical information using both the pitch (note and scale degree) and timing (rhythm and inter-onset-interval) representation described in

[1] As conceptualized as either a Schenkerian foreground (**?**) or defined by **?**

**?**. Future research comparing this model's output using different representations will also contribute to conversations regarding pedagogy in that if one form of data representation mirrors human behavior better than others, it would provide more than evidence in support of the pedagogy of one system over another. How the model represents musical information is the first important parameter value that needs be chosen before running the model and this establishes the *Prior Knowledge*.

### 7.3.2 Contents of the Prior Knowledge

The *Prior Knowledge* consists of a corpus of digitally represented melodies taken to reflect the implicitly understood structural patterns in a musical style that the listener has been exposed to. The logic of representing an individual's prior knowledge follows the assumptions of both the Statistical Learning Hypothesis (SLH) and the Probabilistic Prediction Hypothesis (PPH), both core theoretical assumptions of the Information Dynamic of Music (IDyOM) model of Marcus Pearce (**??**). Using a corpus of melodies to represent an individual's prior knowledge relies on the Statistical Learning Hypothesis which states:

> musical enculturation is a process of implicit statistical learning in which listeners progressively acquire internal models of the statistical and structural regularities present in the musical styles to which they are exposed, over short (e.g., an individual piece of music) and long time scales (e.g., an entire lifetime of listening). p.2 (Pearce, 2018)

The logic here is that the more an individual is exposed musical material, the more they will implicitly understand it which leads the corroborating probabilistic prediction hypothesis which states:

> while listening to new music, an enculturated listener applies models learned via the SLH to generate probabilistic predictions that enable them to organize and process their mental representations of the music and generate culturally appropriate responses. p.2 (Pearce, 2018).

Taken together and then quantified using Shannon information content (**?**), it then becomes possible using the IDyOM framework to have a quantifiable measure that reliably predicts the amount of perceived unexpectedness in a musical melody that can change pending on the musical corpus that the model is trained on. As a model IDyOM has been successful mirroring human behavior in melodies in various styles (**?**), harmony– outperforming (**?**) sensory models of harmony (**?**)–, and is also being developed to handle polyphonic materials (**?**).

Stepping beyond the assumptions of IDyOM, the prior knowledge also needs to have a implicit/explicitly known parameter which indicates whether or not an pattern of music– or n-gram[2] pattern– is explicitly learned. This threshold can be set relative to the entire distribution of all n-grams in the corpus.

### 7.3.3 Modeling Information Content

Having established that the models' first parameters to be decided are the representation of strings and the implicit/explicit threshold, the next decision that has to be made is how the model decides segmentation for the second stage of *Selective Attention*. Although there has been a large amount of work on different ways to segment the musical surface using rule based methods (**????**), which rely on matching a music theorist's intuition with a set of descriptive rules somewhat like the boundary formation rules put forward in *A Generative Theory of Tonal Music*, as noted by Pearce (**?**), rule based models often fail at when applied to music outside the Western art music canon. Additionally, since melodic dictation is an active memory process, rather than a semi-passive process of listening, this model needs to be able to quantify musical information on two conditions. The first is that it must be dependent on prior musical experience. The second is that it should allow for a movable boundary for selective attention so that musical information that is memory can be actively maintained while carrying out another cognitive process, that of notating the melody.

[2] n-grams refer to the amount of musical objects in a string. For example a bi-gram or 2-gram, would be an interval. Tri-grams or 3-grams would consist of two intervals and so on.

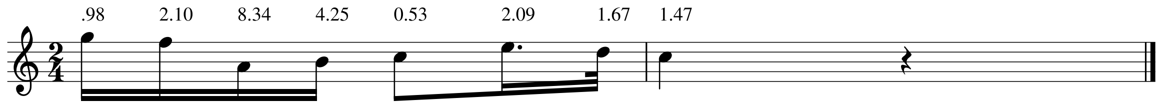


Figure 7.1: Cadential Excerpt from Schubert's Octet in F Major

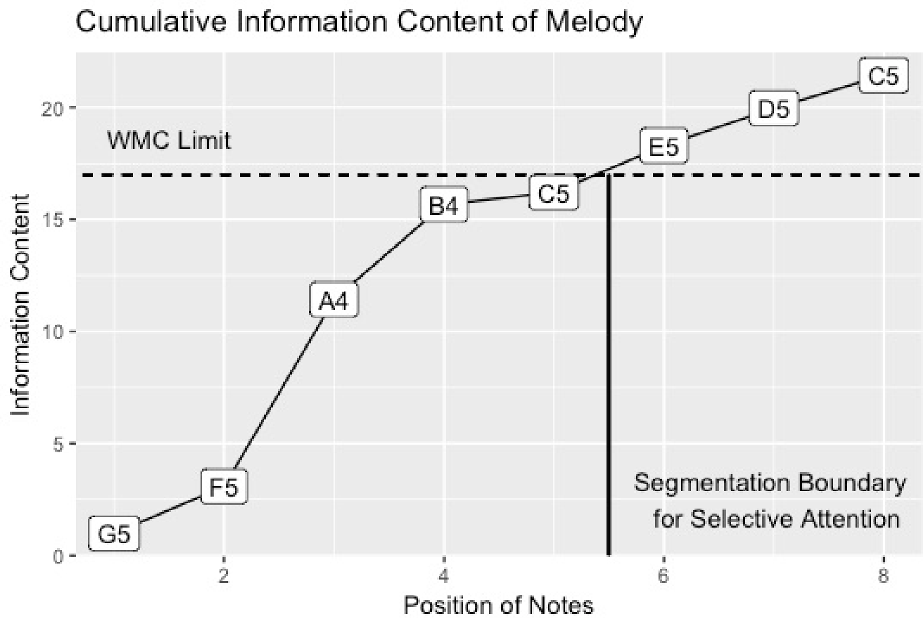


Figure 7.2: Cumulative Information in Schubert Octet Excerpt

In order to create this metric, I rely on IDyOM's use of information content (**?**) which quantifies the information content of melodies based on corpus of materials. For example, when trained against a corpus of melodies, this excerpt in Figure 7.1 from the fourth movement of Schubert's *Octet in F Major* (D.803) lists the information content of the excerpt calculated for each note atop the notation[3] Appearing in Figure 7.2, I plot the cumulative information content of the melody, along with both an arbitrary threshold for the limits of working memory capacity and where the subsequent segmentation boundary for musical material to be put in the *Selective Attention* buffer would be. These values chosen show a small example of how the *Selective Attention* module works. The advantage of operationalizing how an individual hears a melody like this is that melodies with lower information content, derived from an understanding of having more predictable patterns from the corpus, will allow for larger chunks to be put inside of the selective attention buffer. Additionally, individuals with higher working memory capacity would be able to take in more musical information.

It is important to highlight that the notes above the melody here are dependent on what is current in the *Prior Knowledge* module. A corpus of *Prior Knowledge* with less melodies would lead to higher information content measures for each set of notes, while a prior knowledge that has extensive tracking of the patterns would lead to lower information content. This increase in predictive accuracy mathematically reflects the

[3]The following musical examples is taken from **?** reflects a model where IDyOM was configured to predict pitch with an attribute linking melodic pitch interval and chromatic scale degree (pitch and scale degree) using both the short-term and long-term models, the latter trained on 903 folk songs and chorales (data sets 1, 2, and 9 from table 4.1 in (**?**) comprising 50,867 notes.

intuition that those with more listening experience can process greater chunks of musical information.

### 7.3.4 Setting Limits with Transcribe

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

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

### 7.3.5 Pattern Matching

With subset of notes of the melody represented in the attentional buffer, whether or not the melody becomes notated depends on whether or not the melody or string in the buffer can be matched with a string that is explicitly known in the corpus. Mirroring a search pattern akin to Cowan's Embedded Process model (**??**), the individual would search across their long term memory, or *Prior Knowledge* for anything close to or resembling the pattern in the *Selective Attention* buffer. Cowan's model differs from other more module based models of working memory like those of **?** by positing that working memory should be conceptualized as a small window of conscious attention. As an individual directs their attention to concepts represented in their long term memory, they can only spotlight a finite amount of information where categorical information regarding what is in the window of attention not far from retrieval. An example of this bottle necking is given after a formal statement of the model. Using this logic, longer pattern strings n-grams would be less likely to be recalled exactly since they occur less frequently in the prior knowledge.

When searching for a pattern match, the *Transcription* module is at work. If a pattern match that has been moved to *Selective Attention* is immediately found, the contents of *Selective Attention* would be considered to be notated. The model would register that a loop had taken place and document the n-gram match. Of course, finding an immediate pattern match each time is highly unlikely and the model needs to be able to compensate if that happens.

If a pattern is not found in the initial search that is *explicitly* known, one token of the n-gram would be dropped off the string and the search would happen again. This recursive search would happen until an explicit long term memory match is made. Like humans taking melodic dictation, the computer would have the best luck finding patterns that fall within the largest density of a corpus of intervals distribution. Additionally, like students performing a dictation, if a student does not explicitly know an interval, or a 2-gram, the dictation would not be able to be completed. If this happens, both the model and student would have to move on to the next segment via the *Re-entry* function.

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

This type of pattern search is also dependent on the way that the *Prior Knowledge* is represented. In the example here, both pitch and rhythmic information are represented in the string. Since there is probably a very low likelihood of finding an exact match for every n-gram with both pitch and rhythm, this pattern search can happen again with both rhythms and pitch information queried separately. If not found, exact pitch-temporal matches are found and the search is run again on either the pitch or rhythmic information separately; this would be computationally akin to Karpinski's proto-notation that he suggests students use in learning how to take melodic dictation (**?**, p.88). This feature of the model would predict that more efficient dictations would happen when pitch and interval information is dictated simultaneously. Running the model prioritizing the secondary search with either pitch or rhythmic information will provide new insights into practical applications of dictation strategies. Using this separate search feature as an option of the model seems to match with the intuitions strategies that someone dictating a melody might use.

### 7.3.6 Dictation Re-Entry

Upon the successful pattern match of a string, the *Selective Attention* and *Transcription* module would need too then be run again. This process is done via the *Re-entry* function. As noted above, re-entry in the melody could be a highly subjective point of discussion. The model could either re-enter at the last note where the function successfully left off, the note in the melody with the lowest information content, the n-gram most salient in the corpus, or theoretically any other type of way that could be computationally implemented. Entering at the last note not transcribed is logical from a computational standpoint, but this linear approach seems to be at odds with anecdotal experience. Entering at the note with the lowest information content seems to provide a intuitive point of re-entry in that it would then be easier to transcribe. Entering at the most represented n-gram seems to match the most with intuition in that people would want to tackle the easier tasks first, but this rests on the assumption that humans are able to reliably detect the sections of a melody that are easiest to transcribe based on implicitly learned statistical patterns. For example, some people might instead choose to go to the end of a melody after successful transcription of the start of the melody. This might be because this part of the melody is most active in memory due to a recency effect, or it could be that that cadential gestures are more common in being represented in the prior knowledge.

### 7.3.7 Completion

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

### 7.3.8 Model Output

The model then outputs each n-gram transcribed and can be counted as a series with less attempts mapping to an easier transcription. I believe that this lines with many intuitions about the process of melodic dictation. It first creates a linear mapping of attempts to dictate with difficulty of the melody. It relies on a distinction between explicit and implicit statistical knowledge. It is based on the Embedded Process Model from working memory and attention, so is part of a larger generative model, giving more credibility that this *could* be how melodic dictation works.

## 7.4 Formal Model

Below I present the computational model in psudeocode as described in Figure 7.3. First listed are the defined inputs, the functions needed to run the algorithm, and then the sequence the model runs. To aid distinguishing between functions and objects, I put functions in italics and objects in bold. Below the model in Figure 7.4, I provide a brief walk through of one iteration of the model.

### 7.4.1 Computational Model

### 7.4.2 Example

The example above shows one iteration of the model run using the musical example from above using a hypothetical corpus for the pattern matching. Using the model above, the following inputs were defined *a priori*:

- The **Prior Knowledge** is a hypothetical corpus of symbolic strings representing all n-grams of melodies
- The **Threshold** is set to **five** exact matches in the **Prior Knowledge**
- The **WMC** is set at 17
- The **Target Melody** is the Schubert excerpt from above
- The **String Position** object is used to track the position in the dictation
- The **Difficulty** object starts at 0
- The **Dictation** object is `NULL` to begin, and each new n-gram successfully transcribed is annexed to it

Figure 7.4 progresses from left to right over the course of time. The algorithm begins by first running the `listen()` function on the **Target Melody**. First the model checks that there are notes to transcribe; this being the first loop of the model, this is statement will be `FALSE` so the next step is taken. Notes of the **Target Melody** are read in to the **Selective Attention** buffer until the information content of the melody exceeds that of the working memory threshold. This is depicted graphically in the leftmost panel of Figure 7.4. Each note unfolding over time fills up the **Selective Attention** working memory buffer. When the amount of information reaches the perceptual bottleneck– as indicated by the dashed line– the **Selective Attention** buffer stops receiving information. At this point the model will mark where in the melody it stopped taking in new information for later. Here the contents in **Selective Attention** are moved to the `transcribe()` function.

With the contents of **Selective Attention** passed to `transcribe()`, the model adds one to the counter indicating the first search is about to run. Moving to the middle panel of Figure 7.4, the symbol string of notes in the first column are indexed against the **Prior Knowledge**. Only if a five note pattern has appeared more than or equal to five times, as determined by the **Threshold** input, will the corresponding `EXPLICIT` column be `TRUE`. In this case, this pattern has occurred over the threshold of 5 and thus a successful match is found. It is at this step that the search resembles that of Cowan's model of working memory as active attention. The pattern being searched for is compared against a vast amount of information, with cues from the contents of what is in **Selective Attention** grouping similar patterns together. At the neural level, this is most likely a much more complex process, but to show this grouping I note that this search is at least

## Computational Model

Pseudocode Notation

Functions = *italicised*
Objects = **bold**

**Define Inputs**

**priorKnowledge** ← corpus of symbolic strings representing all possible n-grams of melodies
Consists of complex (IDyOM) and simple (pitch and rhythm) representation
**threshold** ← threshold set for **priorKnowledge** that determines which n-grams are explicitly represented
**wmc** ← individual limit on amount of information that can be held in memory
**selectiveAttention** ← buffer used to hold truncated melodies
**targetMelody**← novel melody represented as symbol string with calculated information content
**stringPosition** ← object used to track position in dictation
**difficulty** ← counter used to track number of iterations of model

**dictation** ← segmented string that holds n-grams parsed by model

**Define Functions**

*listen* ← function(**targetMelody**){
1. IF length(targetMelody == 0 { DONE }
2. ELSE{ Read in symbols of target melody until melody information content >= **wmc**
3. Put symbols into **selectiveAttention**
4. **stringPosition** ← floor(selectiveAttention$position)
5. Move contents of **selectiveAttention** to *transcribe* }

*transcribe* ← function(**selectiveAttention**){
1. Current string counter ++
2. Pattern match **selectiveAttention** to corpus where explicit == TRUE
    a. IF(Match == TRUE) { run notateReentry on **selectiveAttention** }
    b. IF(NO match found) { drop 1 token; re-run *transcribe* }
    c. IF(NO 2-gram found) { run separate searches on **priorKnowledge** simple notation}
3. Pattern match **selectiveAttention** to **priorKnowledge** pitch representation where explicit == TRUE
4. Pattern match **selectiveAttention** to **priorKnowledge** rhythm representation where explicit == TRUE
5. If no 2-grams found, run *notateReentry* with noMatch == TRUE

*notateReentry* ← function(**selectiveAttention**, noMatch == FALSE ){
1. IF (noMatch == TRUE) { run *listen* at position **stringPosition** + 1 }
2. ELSE { **dictation** ←← **selectiveAttention**; run *listen* at position **stringPosition** + 1 }

**Run Model**

*listen(***targetMelody)**
*transcribe*()
*notateReentry*()

Figure 7.3: Formal Model

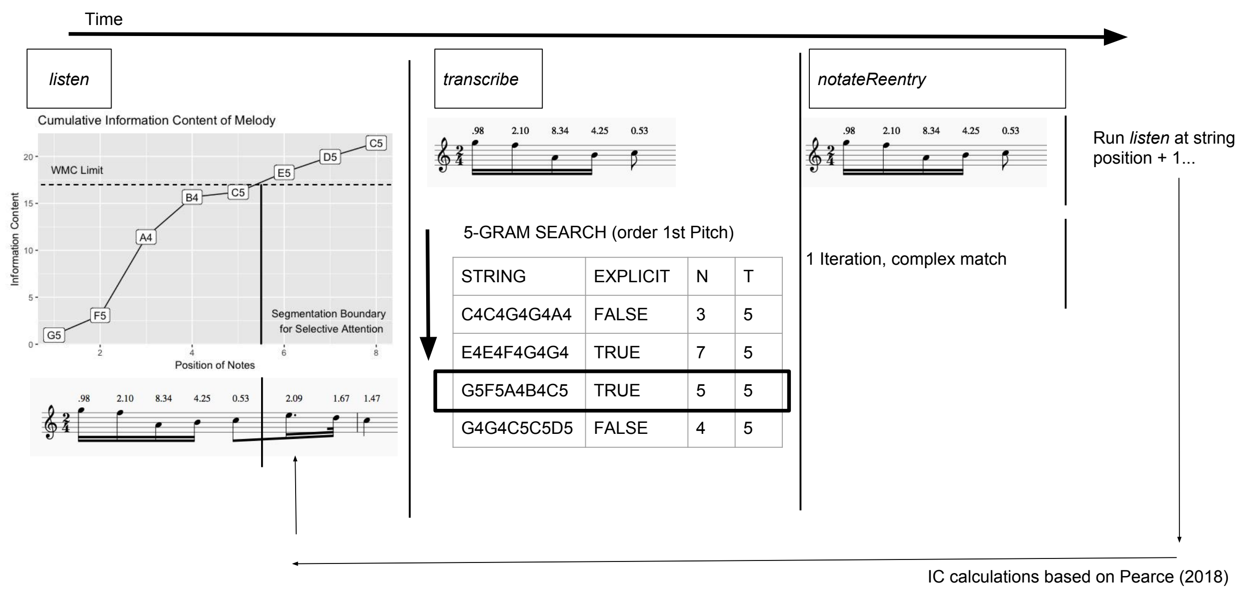


Figure 7.4: Model Example

organized by the first pitch. I assume it would be reasonable that patterns starting on G as $\hat{5}$[4] might happen together. Since this string does have a `TRUE` match with `EXPLICIT`, the contents of **Selective Attention** are considered notated. At this point the model would record the 5-gram, along with the string that it was matched with. the function would then re-run the `listen` function via the `notateReentry()` function at the next point in the melody as tracked by the **String Position** object.

If there were not to have been an exact match, the model would remove one token from the melody and performed the search again on the knowledge of all 4-grams and add one to the **Difficulty** counter. This process would happen recursively until a match is found. If no match is found in either the complex representation, or that of the two rhythm and pitch corpora, the fifth step of `transcribe()` would trigger `notateReentry()` to be run without documenting the n-gram currently being dictated. This would be akin to a student not being able to identify a difficult interval, thus having to restart the melody at a new position. Decisions about re-entry warrant further research and discussion, but this model for the sake of parsimony, assumes linear continuation. As notated in §7.3.5, other modes of re-entry could be incorporated into the model.

This looping process would occur again and again until the entire melody is notated. With each iteration of each n-gram notated, the difficulty counter would increase in relation to the representation of that string in the corpus. This provides an algorithmic implementation of a theorist's intuition that less common n-grams or intervals (2-grams) are going to lead to higher difficulty in dictation. Also worth noting is steps 3 and 4 in the `transcribe()` function are akin to Karpinski's proto-notation. Further research might consider advantages in the order of searching the **Prior Knowledge** corpora.

## 7.5 Conclusions

In this chapter, I presented an explanatory, computational model of melodic dictation. The model combines work from computational musicology and work from cognitive psychology. In addition to being a complete model that explicates every step of the dictation process, the model seems to match phenomenological intuitions as to the process of melodic dictation. Given the current state of the model, it makes predictions

[4]As determined by being calculated against the corpus with both pitch and scale degree information

about the dictation process and can eventually be implemented and tested against human behavioral data to provide evidence in support of its verisimilitude. For example, the model predicts:

- Segments of melodies are likely successfully to be dictated relative to the frequency distribution of their prior knowledge.
- Higher working memory span individuals will be able to dictate bigger chunks of melodies, and thus perform better at dictation
- Using an *atomistic* dictation will result not as effective dictations than attempting to identify larger patterns
- Determining the difficulty of melodies of equal length is predictable from the frequency the melody's cumulative n-gram distribution.
- Some *atonal* melodies will be easier to dictate than tonal melodies if they consist of patterns that are more frequent in a listener's prior knowledge
- Higher exposure to sight-singing results in more explicitly learned patterns, thus the ability to identify larger patterns of music

Although many of these hypotheses might seem intuitive to any instructor that has taught aural skills before, work from this dissertation provides a theory as to why each appears to be true. Future research beyond this dissertation will explore further predictions of this work in more detail. Most importantly from a pedagogical standpoint, the model and underlying theory gives exact language as to how and why melodic dictation works, which can serve as a valuable pedagogical and research contribution.

# Chapter 8

# Reference Log

## 8.1 To Incorporate

- Grutzmaker
- (?) – Margulis Model
- (?) – Specialty jazz background helps in tasks, WMC
- (?)
- (?)
- (?) – Musical Characteristics predict memory
- (?) – Great citation that lots of things change memory, even structural!
- tallaricoStudyThreePhase1974 – Long boring talk on STM, LTM
- Krof on alternatives to melodic dictation
- (?) – Awful experimental design that says people use structual tones
- (?) – Call for experimental, suggestions as to what factors might contribute, use of deductive reasoning, qualitative
- (?) – People need to focus right away, not establish, distractors
- (?) – Showing people visual music does not help much.
- (?) – Listening helps with other things, no best strategy in terms of writing
- (?) – 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
- (?) – 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
- (?) – music ed study with weird stats, has references to follow up on with advantages of pitch systems and people who reccomend things for sight singing
- (?) – Effects of melodies, also how people do it. Interesting that they too effect of melodies, but talka bout things in terms of notes and not in terms of information content. Thought ot have an experiment where the n-grams that are more common are easier to write down. Lots of good charts too.
- (?) – 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.
- (?) – Good references that people are awful at Aural Skills, Also suggestions that people are not that great at transfer, and some stuff to suggest academic abililty is intertwined in all of this. Good reference for when starting to talk about untangling the mess that is aural skills.
- (?) – 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
- (?) – 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

- (**?**) – 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
- (**?**) – 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
- (**?**) – dictation teacher surprised that people don't keep up their dictaiton skills quote

## 8.2 Chapter 3

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

- (**?**) – 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.

Akiva-Kabiri, L., Vecchi, T., Granot, R., Basso, D., and Schön, D. (2009). Memory for Tonal Pitches: A Music-Length Effect Hypothesis. *Annals of the New York Academy of Sciences*, 1169(1):266–269.

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

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

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

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

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

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

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

Bartlett, J. C. and Dowling, W. J. (1980). Recognition of Transposed Melodies: A Key-Distance Effect in Developmental Perspective. page 15.

Beaujean, A. A. (2014). *Latent Variable Modeling Using R: A Step by Step Guide.* Routledge/Taylor & Francis Group, New York. OCLC: ocn868199963.

Berkowitz, S., editor (2011). *A New Approach to Sight Singing.* W.W. Norton, New York, 5th ed edition. OCLC: ocn653123073.

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

Best, H. M. (1992). Music Curricula in the Future. *Arts Education Policy Review*, 94(2):2–7.

Bigand, E., DelbÃ©, C., Poulin-Charronnat, B. A. A., Leman, M., and Tillmann, B. (2014). Empirical evidence for musical syntax processing? Computer simulations reveal the contribution of auditory short-term memory. *Frontiers in Systems Neuroscience*, 8.

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

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

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

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

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

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

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

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

Butler, D. and Lochstampfor, M. Bridges Unbuilt: Comparing the Literatures of Music Cognition and Aural Training. *Indiana Theory Review*, 14(2):1–17.

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

Cattell, R. (1971). *Abilities: Their Growth, Structure, and Action*. Houghton Mifflin, Boston, MA.

Chenette, T. K. (2019). Reframing Aural Skills Instruction Based On Research in Working Memory. *Journal for Music Theory Pedagogy*, page 18.

Cleland, K. D. and Dobrea-Grindahl, M. (2010). *Developing Musicianship through Aural Skills: A Holisitic Approach to Sight Singing and Ear Training*. Routledge, New York. OCLC: ocn299711396.

Cohen, A. J., Cuddy, L. L., and Mewhort, D. J. K. (1977). Recognition of transposed tone sequences. *The Journal of the Acoustical Society of America*, 61(S1):S87–S88.

Condit-Schultz, N. (2019). Deconstructing the nPVI: A Methodological Critique of the Normalized Pairwise Variability Index as Applied to Music. *Music Perception: An Interdisciplinary Journal*, 36(3):300–313.

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. and Schellenberg, E. G. (2015). Predicting who takes music lessons: Parent and child characteristics. *Frontiers in Psychology*, 6.

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

Corrigall, K. A. and Trainor, L. J. (2011). Associations Between Length of Music Training and Reading Skills in Children. *Music Perception: An Interdisciplinary Journal*, 29(2):147–155.

Coumel, M., Christiner, M., and Reiterer, S. M. (2019). Second Language Accent Faking Ability Depends on Musical Abilities, Not on Working Memory. *Frontiers in Psychology*, 10.

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

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

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

Cuddy, L. L. and Lyons, H. I. (1981). Musical pattern recognition: A comparison of listening to and studying tonal structures and tonal ambiguities. *Psychomusicology: A Journal of Research in Music Cognition*, 1(2):15–33.

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

Daniele, J. R. and Patel, A. D. (2004). THE INTERPLAY OF LINGUISTIC AND HISTORICAL INFLUENCES ON MUSICAL RHYTHM IN DIFFERENT CULTURES. page 5.

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

Degé, F., Kubicek, C., and Schwarzer, G. (2011). Music Lessons and Intelligence: A Relation Mediated by Executive Functions. *Music Perception: An Interdisciplinary Journal*, 29(2):195–201.

Demorest, S. M., Morrison, S. J., Jungbluth, D., and Beken, M. N. (2008). Lost in Translation: An Enculturation Effect in Music Memory Performance. *Music Perception: An Interdisciplinary Journal*, 25(3):213–223.

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

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

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

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

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

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

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

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

Ettlinger, M., Margulis, E. H., and Wong, P. C. M. (2011). Implicit Memory in Music and Language. *Frontiers in Psychology*, 2.

Field, A. P., Miles, J., and Field, Z. (2012). *Discovering Statistics Using R*. Sage, London ; Thousand Oaks, Calif. OCLC: ocn760970657.

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 30573561774514.

Francois, C., Chobert, J., Besson, M., and Schon, D. (2013). Music Training for the Development of Speech Segmentation. *Cerebral Cortex*, 23(9):2038–2043.

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.

Gibson, C., Folley, B. S., and Park, S. (2009). Enhanced divergent thinking and creativity in musicians: A behavioral and near-infrared spectroscopy study. *Brain and Cognition*, 69(1):162–169.

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.

Grabe, E. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. *Papers in laboratory phonology*, page 16.

Halpern, A., Bartlett, J., and Dowling, W. (1995). Aging and Experience in the Recognition of Musical Transpositions. *Psychology and Aging*, 10(3):325–342.

Halpern, A. R. and Bartlett, J. C. (2010). Memory for Melodies. In Riess Jones, M., Fay, R. R., and Popper, A. N., editors, *Music Perception*, volume 36, pages 233–258. Springer New York, New York, NY.

Halpern, A. R. and Müllensiefen, D. (2008). Effects of timbre and tempo change on memory for music. *Quarterly Journal of Experimental Psychology*, 61(9):1371–1384.

Hambrick, D. Z., Burgoyne, A. P., and Oswald, F. L. (2019). Domain-General Models of Expertise: The Role of Cognitive Ability. In Ward, P., Maarten Schraagen, J., Gore, J., and Roth, E., editors, *The Oxford Handbook of Expertise*. Oxford University Press.

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

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

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

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

Harrison, P. M. . C. and Pearce, M. T. (2018). Dissociating sensory and cognitive theories of harmony perception through computational modeling.

Harrison, P. M. C., Collins, T., and Müllensiefen, D. (2017). Applying modern psychometric techniques to melodic discrimination testing: Item response theory, computerised adaptive testing, and automatic item generation. *Scientific Reports*, 7(1).

Hick, W. E. (1952). On the Rate of Gain of Information. *Quarterly Journal of Experimental Psychology*, 4(1):11–26.

Hille, K., Gust, K., Bitz, U., and Kammer, T. (2011). Associations between music education, intelligence, and spelling ability in elementary school. *Advances in Cognitive Psychology*, 7(-1):1–6.

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

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

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

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

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, 45(3):188–196.

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

Jamieson, R. K. and Mewhort, D. J. K. (2009). Applying an exemplar model to the serial reaction-time task: Anticipating from experience. *Quarterly Journal of Experimental Psychology*, 62(9):1757–1783.

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.

Karpinski, G. S. (2007). *Manual for Ear Training and Sight Singing.* Norton, New York.

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.

Koo, T. K. and Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2):155–163.

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

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

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

Korenman, L. M. and Peynircioğlu, Z. F. (2004). The Role of Familiarity in Episodic Memory and Metamemory for Music. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4):917–922.

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

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

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

Lartillot, O. and Toiviainen, P. (2007). A Matlab Toolbox for Musical Feature Extraction from Audio. *International Conference on Digigial Audio Effects*, pages 237–244.

Lerdahl, F. and Jackendoff, R. (1986). *A Generative Theory of Tonal Music.* MIT Press, Cambridge.

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

Levitin, D. J. and Tirovolas, A. K. (2009). Current Advances in the Cognitive Neuroscience of Music. *Annals of the New York Academy of Sciences*, 1156(1):211–231.

Li, D., Cowan, N., and Saults, J. S. (2013). Estimating working memory capacity for lists of nonverbal sounds. *Attention, Perception, & Psychophysics*, 75(1):145–160.

Lomax, A. (1977). *Cantometrics: An Approach to the Anthropology of Music: Audiocassettes and a Handbook.* University of California Press, Berkle, California.

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.

Margulis, E. H. (2014). *On Repeat: How Music Plays the Mind*. Oxford University Press, Oxford.

McAuley, J., Stevens, C., and Humphreys, M. S. (2004). Play it again: Did this melody occur more frequently or was it heard more recently? The role of stimulus familiarity in episodic recognition of music. *Acta Psychologica*, 116(1):93–108.

McDermott, J. H., Schultz, A. F., Undurraga, E. A., and Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature*, 535(7613):547–550.

McFee, B., Raffel, C., Liang, D., Ellis, D. P. W., McVicar, M., Battenberg, E., and Nieto, O. (2015). Librosa: Audio and Music Signal Analysis in Python. page 7.

Mehr, S. A., Schachner, A., Katz, R. C., and Spelke, E. S. (2013). Two Randomized Trials Provide No Consistent Evidence for Nonmusical Cognitive Benefits of Brief Preschool Music Enrichment. *PLoS ONE*, 8(12):e82007.

Meinz, E. J. and Hambrick, D. Z. (2010). Deliberate Practice Is Necessary but Not Sufficient to Explain Individual Differences in Piano Sight-Reading Skill: The Role of Working Memory Capacity. *Psychological Science*, 21(7):914–919.

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

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

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

Miyake, A. and Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, 21(1):8–14.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology*, 41(1):49–100.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. a. L., and Besson, M. (2009). Musical Training Influences Linguistic Abilities in 8-Year-Old Children: More Evidence for Brain Plasticity. *Cerebral Cortex*, 19(3):712–723.

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

Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLoS ONE*, 9(2):e89642.

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

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

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

Mungan, Peynirvioglu, and Halpern (2011). Levels-of-Processing Effects on "Remember" Responses in Recognition for Familiar and Unfamiliar Tunes. *The American Journal of Psychology*, 124(1):37.

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.

Okada, B. M. and Slevc, L. R. (2018). Individual differences in musical training and executive functions: A latent variable approach. *Memory & Cognition*.

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

Ottman, R. W. and Rogers, N. (2014). *Music for Sight Singing.* Pearson, Upper Saddle River, NJ, 9th ed edition.

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.

Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., and Kraus, N. (2011). Musical Experience and the Aging Auditory System: Implications for Cognitive Abilities and Hearing Speech in Noise. *PLoS ONE*, 6(5):e18082.

Patel, A. D. and Daniele, J. R. (2003). Stress-Timed vs. Syllable-Timed Music? A Comment on Huron and Ollen (2003). *Music Perception*, 21(2):273–276.

Pearce, M. (2005). *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition.* PhD thesis, City University of London., Department of Computer Science.

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

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

Pearl, J. and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect.* Basic Books, New York, first edition edition.

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. (2006). The nature of music from a biological perspective. *Cognition*, 100(1):1–32.

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

Perruchet, P. and Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10(5):233–238.

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

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

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

Raven, J. (1994). *Manual for Raven's Progressive Matrices and Mill Hill Vocabulary Scales.*

Remillard, G. and Clark, J. M. (2001). Implicit learning of first-, second-, and third-order transition probabilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(2):483–498.

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

Rohrmeier, M. and Rebuschat, P. (2012). Implicit Learning and Acquisition of Music. *Topics in Cognitive Science*, 4(4):525–553.

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

Rosseel, Y. (2012). Lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2).

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.

Sauve, S. (2017). *Prediction in Polyphony: Modelling Auditory Scene Analysis*. PhD thesis, Queen Mary, University of London, Centre for Digital Music.

Savage, P. E., Brown, S., Sakai, E., and Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences*, 112(29):8987–8992.

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

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

Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. *Journal of Educational Psychology*, 98(2):457–468.

Schellenberg, E. G. (2011). Examining the association between music lessons and intelligence: Music lessons and intelligence. *British Journal of Psychology*, 102(3):283–302.

Schellenberg, E. G. (2015). Music training and speech perception: A gene-environment interaction: Music training and speech perception. *Annals of the New York Academy of Sciences*, 1337(1):170–177.

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

Schellenberg, E. G. and Mankarious, M. (2012). Music training and emotion comprehension in childhood. *Emotion*, 12(5):887–891.

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.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423.

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.

Steinbeck, W. (1982). *Struktur Und Ähnlichkeit: Methoden Automatiserter Melodieanalyse.* Bärenreiter, Kallel.

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

Strait, D. L., Parbery-Clark, A., Hittner, E., and Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language*, 123(3):191–201.

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

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

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

Team, R. C. (2015). R: A language and environment for statistical computing.

Temperley, D. (2004). *The Cognition of Basic Musical Structures.* MIT Press, Cambridge.

Thurstone, L. L. (1938). *Primary Mental Abilities.* University of Chicago Press, Chicago.

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

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., and Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, 17(6):635–654.

VanHandel, L. and Song, T. (2010). The Role of Meter in Compositional Style in 19th Century French and German Art Song. *Journal of New Music Research*, 39(1):1–11.

VanHandel, L., Wakefield, J., and Wilkins, W. K. (2011). Towards the role of working memory in pitch processing in language and music. In Rebuschat, P., Rohmeier, M., Hawkins, J. A., and Cross, I., editors, *Language and Music as Cognitive Systems*, pages 302–309. Oxford University Press.

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

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

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

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

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

Wright, S. (1934). The method of path coefficents. *The Annals of Mathematical Statistics*, 5(3):161–215.