The Society for Music Theory
2019 SMT-40 Dissertation Fellowship Application Form

Attach a completed copy of this form and PDFs of documents (2)–(4) indicated in the application guidelines to an email addressed to Ellie Hisama, the Chair of the SMT-40 Dissertation Fellowships Committee, at eh2252@columbia.edu. Include your last and first names in the names of the attached files (last_first_document).

Dissertation committee chairs should e-mail their confidential letters directly to the Chair at eh2252@columbia.edu.

Deadline for receipt of application emails and attachments: OCTOBER 26, 2018.

Name	David John Baker
Mailing address	2131 N 72nd Street, Wauwatosa, WI, 53213
Email address	davidjohnbaker1@gmail.com
Academic institution	Louisiana State University
Chair of dissertation committee	Dr. Daniel Shanahan
Chair's email address	shanahan.37@osu.edu
Expected date for award of Ph.D.	May 2019

By submitting this application, I certify that by October 26, 2018 I will have completed all Ph.D. requirements other than the dissertation.

David John Baker

October 23, 2018

Contact Information
davidjohnbaker1@gmail.com
DavidJohnBaker
david_john_baker
+1 414 736 7948

Online Platforms
ResearchGate
Github
My Website

EDUCATION

Ph.D, Music Theory, ABD — Defense: May, 2019
Cognitive and Brain Sciences Minor
Louisiana State University, Baton Rouge, Louisiana, USA
Dissertation – Modeling Melodic Dictation

MSc., Music, Mind and Brain — September, 2014
Goldsmiths, University of London, England
Thesis – Salience and Memorability of Leitmotives in the Music of Richard Wagner

B.M., Instrumental Performance — May, 2012
Baldwin Wallace University, Conservatory of Music, Berea, OH, USA

PUBLICATIONS

Journal Articles

Baker, D., Ventura, J., Calamia, M., Shanahan, D., Elliott, E. (forthcoming). Examining Musical Sophistication: Replication of the Goldsmiths Musical Sophistication Index. *Musicae Scientiae.*(Accepted October 2018)

Akkermans, J., Schapiro, R., Müllensiefen, D., Frieler, K., Jakubowski, K., Busch, V., Fischinger, T., Lothwesen, K., Schlemmer, K., Shanahan, D., & **Baker, D.** (forthcoming) Decoding emotions in expressive music performances: A multi-lab replication and extension study. *Cognition and Emotion.* (Accepted October 2018)

Baker, D. and Müllensiefen D. (2017). Perception of Leitmoitves in Richard Wagner's *Der Ring des Nibelungen.* Frontiers in Psychology: Cognition, Special Issue on Bridging Music Informatics with Music Cognition. [Article]

Book Chapters

Baker, D. and Shanahan, D. (forthcoming). Examining Fixed and Relative Measurements of Similarity through Jazz Melodies. In World Scientific (Ed.) *Scholarly Approaches to Mathematical Music Theory: Algebra and Combinatorics, Geometry, Topology, and Graph Theory, Discrete Fourier Transform and Distance Measures for Music. Forthcoming* [Chapter]

Müllensiefen, D., **Baker, D.**, Rhodes, C,. Crawford, T,. Dreyfus, L. (2016). Recognition of leitmotives in Richard Wagner's music: chroma distance and listener expertise. In Springer (Ed.) *Analysis of Large and Complex Data.* [Chapter]

Select Conference Proceedings/Talks

Baker, D., Elliott, E., Shanahan, D., Ventura, J. Monzingo, E., Ritter, B., & Young, C. (2018) Explaining Objective and Subjective Aspects of Musical Sophistication: Insights from General Fluid Intelligence and Working Memory. Proceedings of 2018 International Conference on Music Perception and Cognition.

Shanahan, D., Elliott, E., **Baker, D.**, Ventura, J., Monzingo, E., Holt, H., & Keller, H. (2018) Dissecting the Effects of Working Memory, General Fluid Intelligence, and Socio-Economic Status on Musical Sophistication. Proceedings of 2018 International Conference on Music Perception and Cognition.

Baker, D., Monzingo, E., & Shanahan, D. (2018) Modeling Aural Skills Dictation. Proceedings of 2018 International Conference on Music Perception and Cognition.

Monzingo, E., **Baker, D.**, & Shanahan. D. (2018) The Relationships Between Genre Preference, Aural Skills, and Tonal Working Memory. Proceedings of 2018 International Conference on Music Perception and Cognition.

Baker, D. (2018). Who counts as a Musician?, Music and Public Discourse, Columbia, South Carolina. [Video]

Reviews

Baker, D., (2018) "Coughing and Clapping", in *Empirical Musicology Review* , Vol 12, No. 5. [Book Review]

Baker, D., (2015) "Unlocking the Mysteries of Your Brain, Dr. Daniel Levitin Public Lecture", in *Psychomusicology: Music, Mind, and Brain*, Vol 25, No.4, pp. 455-456. [Public Lecture Review]

Manuscripts in Preparation

Baker, D. (2019). Modeling Melodic Dictation. *Doctoral Dissertation: LSU University.*

Baker, D. (May, 2018). Understanding Melodic Dictation via Experimental Methods. In *Companion to Aural Training in Music Education.* Routledge *Chapter Proposal Accepted* [Chapter]

Other Publications

Baker, D (2018) R For Psychologists Handbook. In Progress Current Version of Project Available Here

Müllensiefen, D., **Baker, D**. (2015) Music in Advertising: Testing What Works. Audio Branding Academy Yearbook 2015.[Article in Yearbook]

INVITED TALKS

Baker, D. (November, 2018) Modeling Melodic Dictation. Research Seminar: Computing & Communications Department, The Open University. Milton Keynes, England.

TEACHING EXPERIENCE

Graduate Multivariate Statistics, Teaching Assistant
LSU Psychology Department, Baton Rouge, Louisiana — Spring, 2018

Intermediate Statistics, Teaching Assistant
LSU Psychology Department, Baton Rouge, Louisiana — Autumn, 2017

Foundations of Music Study, Instructor of Record
LSU Music Department, Baton Rouge, Louisiana — Autumn, 2016

Aural Skills IV, Instructor of Record
LSU Music Department, Baton Rouge, Louisiana
Semesters Taught — Spring 2015, Spring 2016

Aural Skills III, Instructor of Record
LSU Music Department, Baton Rouge, Louisiana
Semesters Taught — Autumn 2015, Autumn 2016

VOLUNTEER WORK

Research and Evaluation Residential Voluntary Worker — June 2018 – March 2019
London, England

- Worked at Toynbee Hall and used a Participatory Action Research (PAR) to train people from local communities to co-designed and produced research using quantitative and qualitative tools
- Topics researched included financial capability, mental health support for advice users, private renting, poverty premium, and support for young people.

PROFESSIONAL WORKSHOPS

Theorizing Categorically: Film Music and Beyond with Scott Murphy. Graduate Student Workshop, South Central Society for Music Theory, Hattiesburg, MS, 2018.

Film Music: Cognition to Interpretation with Dr. Juan Chattah. Graduate Student Workshop, Music Theory South East, Columbia, SC, 2018.

Schubert: Text and Music with Dr. Jeffery Perry. Graduate Student Workshop, South Central Society for Music Theory, Memphis, TN, 2017.

Music Informatics Group Workshop on Computational Music Analysis. Green College, University of British Columbia, Canada. November, 2016.

Introduction to Empirical Musicology Summer School, with Dr. David Huron . Ohio State University, Ohio, 2012.

MEMBERSHIPS

Society of Music Theory, Graduate Student Member
South Central Society of Music Theory, Graduate Student Member
Society of Music Perception and Cognition, Student Member

Modeling Melodic Dictation

SMT40-Prospectus

David John Baker

October 26th, 2018

Motivation

Melodic dictation is the process in which an individual hears a melody, retains it in memory, 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. This is evident from the fact that most aural skills textbooks with content devoted to honing one's listening abilities have sections devoted to learning how to take melodic dictation (Karpinski 2000). Additionally, any school accredited by the National Association of Schools of Music in North America requires students to learn this skill ("National Association of Schools of Music Handbook" 2018, sec. VIII.6.B.2.A). Yet, despite this ubiquity in the pedagogy of melodic dictation, exactly *how* this process works is not understood: Music theorists do not have an explanatory theory of melodic dictation.

The lack of knowledge regarding melodic dictation is alarming given the degree music theorists are engaged with the teaching and assessing of this ability. As a community, a more systematic understanding of *how* people learn melodies is not only important from an pedagogical point of view, but understanding how people learn and perceive melodies is at the locus of research central to music theory, music education, as well as music cognition. While there have been repeated calls throughout the past few decades to synthesize these disparate literatures (Butler 1997; Karpinski 2000; Klonoski 2000), the literature is sparse in relation to how frequent melodic dictation appears as part of our curricula.[1] Reviewing the current state of research on melodic dictation highlights the need for the music theory community to better understand melodic dictation and the literature that surrounds it.

Much of the fundamental work on melodic dictation was synthesized via the work of Gary Karpinski. Originally appearing in an article from 1990, and then later the focus of of the third chapter of *Aural Skills Acqusition*, Karpinski proposes a four step model that describes an idealized process of melodic dictation (Karpinski 2000, 1990). The four steps include:

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

and are conceptualized as a looping process that is done over each chunk of musical material that the listener focuses on via a process of extractive listening. His flowchart of the process is reproduced below in Figure 1.

As a pedagogical tool, Karpinski's model distills a complicated and almost esoteric process into a manageable system that benefits both students as well as aural skills pedagogues. Karpinski's model describes the process of melodic dictation but his model makes no claims as to *how* the process happens. Though not the original intention of the model, this four step model lacks robustness in that it is agnostic to both differences at the individual level, as well as for differences in melodic material.

For example, Karpinski suggests an example for discussion based on his idealized process and claims that listeners with "few to no chunking skills" should be able to dictate a melody of twelve to twenty notes long with two passes of his flowchart. While this provides an approximate rule of thumb as to what can be expected of students, these suggestions are generated from a fixed system and are not flexible to individuals

[1] Paney (2016) notes that since 2000 only four studies explicitly examining melodic dictation have been published in music pedagogy journals.

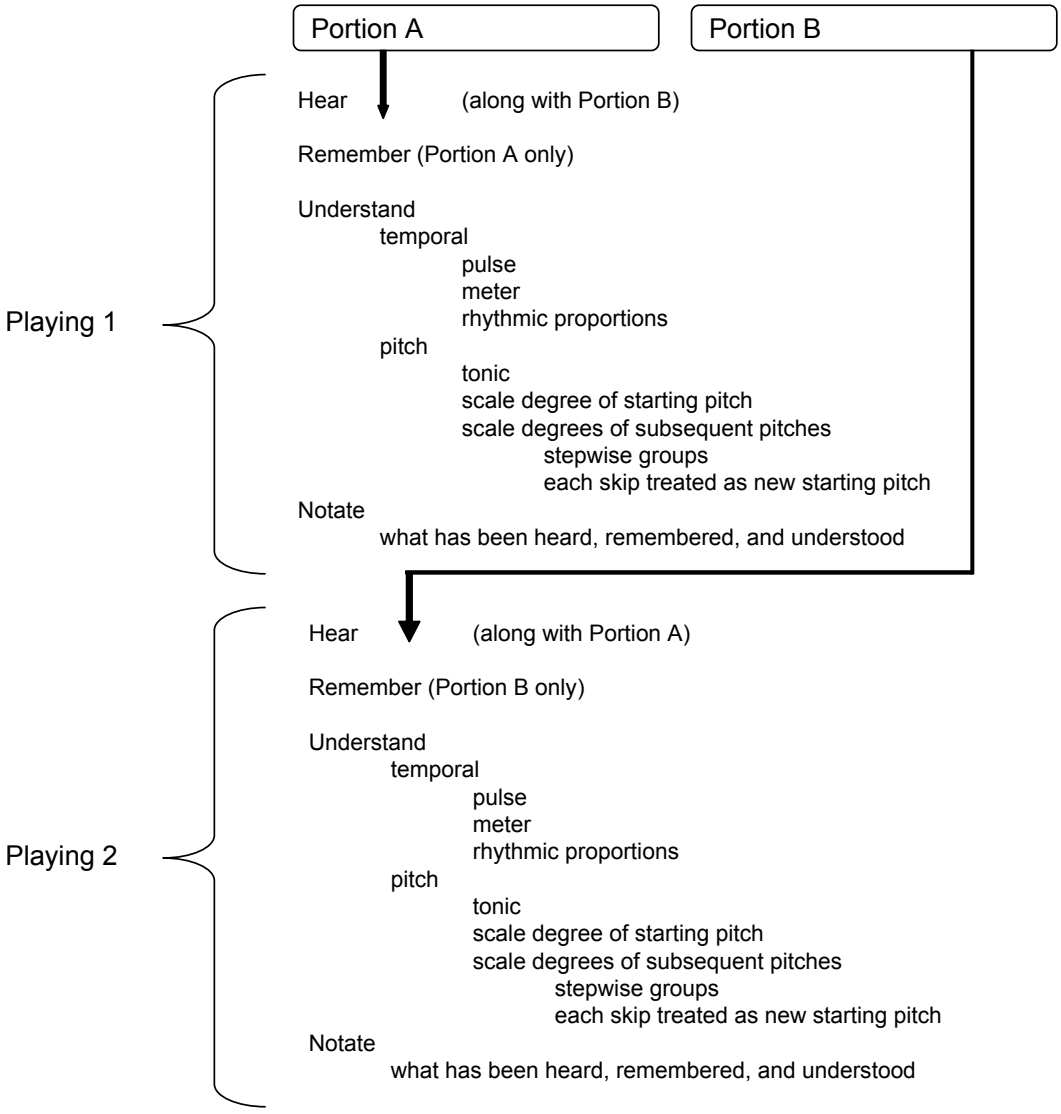

Figure 1: Karpinski Idealized Flow Chart of Melodic Dictation

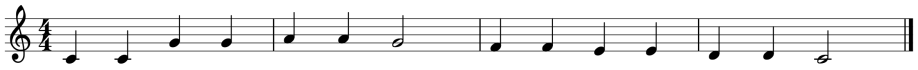

Figure 2: Figure 2: Melodies of Equal Length: Tonal

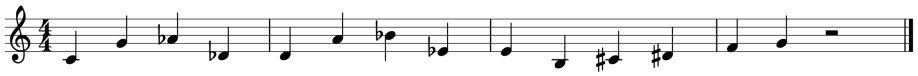

Figure 3: Figure 3: Melodies of Equal Length: Atonal

with different experience levels.[2] To give an example, both melodies from Figure 2 and 3 are 14 notes in length, yet the strategies and thus perceived difficulty of dictating each melody would be presumably different for individuals with different musical training backgrounds. Not only will an individual's prior experience affect this process, but presumably the melodic material will as well.

In this dissertation, I continue the line of research begun with Gary Karpinski's four step model of melodic dictation; I do this by first exploring how both individual experiences and musical material can affect melodic dictation separately, examine how these factors interact, then finally posit an explanatory computational model of melodic dictation.

The past 20 years of research since *Aural Skills Acquisition* have highlighted the importance of taking into account both ability to understand how individual differences play a role in music perception tasks as well as an ability to quantify and operationalize differences in melodies that reflect a theorists' intuitive understanding of melodic material.

In order to organize and then reflect on the vast amount of factors that could contribute to how a person performs in melodic dictation I propose a taxonomy– seen in Figure 4– of parameters with both individual (e.g. cognitive and environmental) and musical (e.g. structural and experimental) parameters from approximately the last two decades that should be further explored when looking at melodic dictation in order to move towards an explanatory theory of melodic dictation.

Using this taxonomy as a guide, I investigate factors thought to contribute to tasks of melodic dictation using a diverse methodological toolbox which borrows techniques ranging from cognitive psychology, to computational musicology, and music theory.

[2]Karpinski does in fact note that as listeners develop more varied skills they can "entertain some significant deviations from such a process," p.103 and gives relevant examples, but these deviations are not formalized which is most likely due to the multiplicity of permutations that may exist.

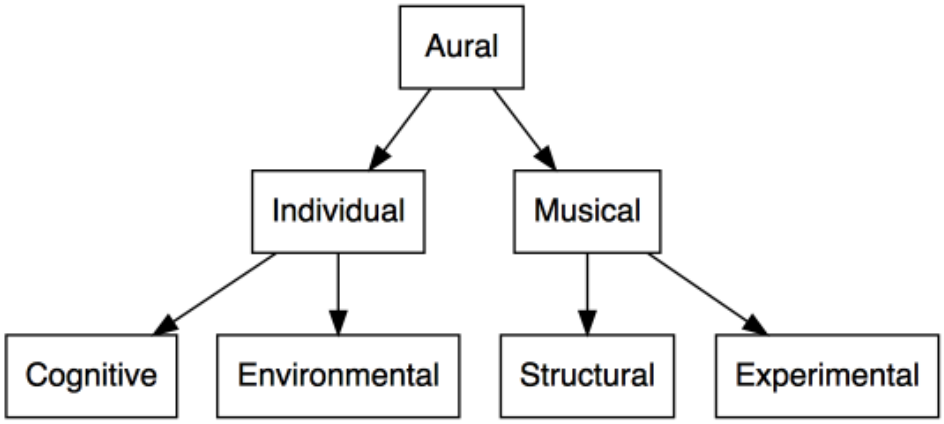

Figure 4: Taxonomy of Factors Relating to Melodic Dictation

Present Research Question

This dissertation takes an interdisciplinary approach to create an explanatory theory of melodic dictation. In order to do this, I synthesize and utilize work from music theory, music education, and music cognition. Since the goal of this research is to improve the music theorists ability to be effective pedagogues, I frame my research question in relation to answering questions as they relate to undergraduate students.[3] Specifically, I write five chapters that respectively investigate the following five questions:

1. To what degree does a student's attention and cognitive ability determine how well they can perform melodic memory tasks?
2. How can computationally abstracted melodic features help determine the difficulty of dictation melody in line with expert intuition?
3. Can patterns from a corpus of sight singing melodies serve as representations of musical structures that students know both implicitly and explicitly?
4. If conceptualized as an experiment, is it possible to predict how a student does on a melodic dictation task when both individual and musical features are accounted for? What aspect is the most predictive of an individual's ability? What problems arise in operationalizing a melodic dictation exercise?
5. Is it possible to posit a computational, explanatory model of melodic dictation that mirrors the phenomenological decision process individuals engage with when performing melodic dictation? If so, how can this be used to help teach this process?

With each of these five questions serving as the basis for each of the five chapters of original content, I next detail what I will use for each of these methodologies in detail.

[3]Although pedagogically focused, the research here also has more domain general applications and relevance.

Outline Research Methods

In order to investigate my first research question, I analyze and interpret data from a large scale experiment conducted over the last year in collaboration with the Louisiana State University's cognitive psychology program. The experiment uses a large sample of (N = 470) students who took participated in a multi part experiment where we took measures of cognitive ability using multiple measures of for both general fluid intelligence and working memory capacity, measures of musical background via subjective self report, and then also an objective test of musical memory via a melodic memory test. Given the complexity of variables at play, I analyze the data using structural equation modeling– a statistical technique developed in order to parse out causal relationships within covariance structures– and find evidence that working memory plays a large role in tasks of musical perception. These results corroborate earlier theoretical claims regarding the need to investigate musical working memory (Berz 1995).

I then introduce a novel corpus in the third chapter that will consist of over 600 sight singing melodies. Using recent advancements in computational musicology, I show how there are statistical norms present in this corpus and suggest that there is a link between using these melodies as an individual's implicit knowledge of a musical structure. In the fourth chapter I demonstrate how computationally extracted features can serve as a quantifiable proxy for a pedagogues' intuition and propose application for future research. Following earlier claims by Meyer (1956) and later computationally driven research by Pearce (2018) I note how information derived from these corpus studies can be incorporated in experiments investigating melodic dictation. This corpus can also be utilized by other researchers who wish to investigate corpus level claims.

Synthesizing the assumptions and findings from the previous three chapters, I combine the work on individual differences and melodic differences in order to operationalize every factor of relevance discussed in earlier chapters in order to conduct an experiment. Here I reflect on the choices needed to be made in this process and talk about implications and limitations of measuring melodic dictation in an experimental setting. Using these experimental methods allows the researcher to determine what happens when the aforementioned musical materials are manipulated in an ecologically valid setting. Findings from this chapter corroborate earlier findings that musical features do in fact play a large role in an individual's ability to perform melodic dictation(Ortmann 1933; Pembrook 1986; Taylor and Pembrook 1983). In detailing the methodologies used in this chapter, I also put forward a complete and free set of software and analysis method so future researchers looking to explore this question on their own are not limited in terms of technical accessibility.

Finally, in the last chapter I put forward a computational model of melodic dictation that explicitly details each step of the melodic dictation process. This Bayesian inspired model takes account of a listener's prior knowledge, explicit understanding of musical material, and then allows a melody to be computationally dictated in order to result in an difficulty score based on previous knowledge, as well as measures thought to be influential like working memory capacity from above. In formalizing this as a computational model, I account for both musical and individual factors relevant to the melodic dictation process, thus providing a new pedagogical and research framework to situate work on melodic dictation.

Significance of Project to Music Theory

Exploring melodic dictation using interdisciplinary approaches allows the music research community to have a more comprehensive understanding of melodic dictation. Firstly, and most importantly and as stated above, there is not an explanatory model of melodic dictation. Discussed in more detail in the attached sample chapter, much of the current work on melodic dictation is supported using descriptive models that only assert what will happen in response to changing one or two aspects of a melodic dictation. Contrary to a descriptive model like Karpinski's Four Step Model, an explanatory model not only posits *what* will happen, but hypothesizes the inner workings of the process and asserts both *how* and *why*. Foreshadowing this included chapter, I combine recent advances in computational musicology via the work of Marcus Pearce's IDyOM model (Pearce 2018)– essentially a computational implementation of theoretical claims made by Meyer (1956)– and Nelson Cowan's Embedded Process model (Cowan 1988, 2010) in order to put forward full, explanatory model of melodic dictation.

Given its importance to the responsibilities of most music theorists, a working theory of how melodic dictation will further our pedagogical understandings. Additionally by putting forward an explanatory theory in the form of a computational model that is informed by experimental evidence, theorists are able to discuss the melodic dictation process with more exact language. Though initially conceptualized as a model for melodic dictation, the model might also have additional predictive power due to its basis in a domain general memory mechanism and could demonstrate how insights from music theory could be used to help explain how the mind works

In terms of novel research findings, this dissertation provides new evidence for some earlier theoretical claims. As mentioned above, work from this dissertation suggests that working memory does in fact play a large role in questions of melodic perception, thus corroborating earlier claims by Berz (1995). This dissertation also puts forward a new corpus of over 600 melodies that can be used by the computational musicology community. Additionally, having written software using open source libraries, new avenues of research can be now open to other researchers interested in setting up experiments investigating melodic dictation with little to no programming experience. Finally, this dissertation posits a new, and explanatory model of melodic dictation based on work from cognitive psychology and computational musicology. In addition to the rationale above, the music theory pedagogy community might be able to more effectively teach melodic dictation. By having a step-by-step model detailing the exact processes used, it will become easier to discuss issues related to melodic dictation like the efficacy of one sight singing system over another, how to best invest time practicing skills related to melodic dictation, and what can be reasonably expected in the assessing of students on this complex ability.

The model also puts forward a series of hypotheses that could be falsified in order to investigate the model's verisimilitude. Among other hypotheses 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*[4] 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.

As stated above, given the ubiquity of melodic dictation in Schools of Music and our responsibility as music theory pedagogues to be able to explain what we do, it is important to have a theory of melodic dictation that is explanatory so students can be correctly assessed on this exercise. As a community it is important to know the degree to which abilities can be learned and how much ability derives from pre-existing individual differences so that we as pedagogues can cater to the diversity of students that we teach. This is especially important given current conversations regarding the content what should be considered central to the curricula for an accredited School of Music education. Throughout the dissertation–though not a central thesis– I argue for a more modular, polymorphic conception of musicianship based on my work on melodic dictation. Discussed mostly in the first chapter literature review and then touched on when relevant, I deconstruct concepts like musicianship and musical sophistication by paralleling recent arguments that deconstruct the notion of a general intelligence factor in psychological research. I argue that conceptualizing musicianship as a set of related, though not unified abilities, we as pedagogues can reconceptualize how we teach and think

[4] A term used by Karpinski to describe the process of trying to hear a melody by listening interval by interval

about the processes that enable or musical choices and be more effective teachers. This argument will appear as part of a forthcoming article that I authored in *Musicae Scientiae*.

Current Progress

As of October 26th, three chapters of the dissertation have been completed as drafts. The completed chapters include the literature review, the fifth chapter on experimental method, as well as the attached chapter that describes the computational model. The second chapter currently exists as a conference proceedings paper from the International Conference on Music Perception and Cognition 15 and I am currently in the process of changing the language and expanding on all of the figures and analyses used in the shorter version of the paper. I am still in the process of encoding the melodies and have finished encoding one third of the melodies in the corpus. The chapters on using computational tools as well as describing the corpus have been outlined, but have not been completed.

Research that resulted from earlier work investigating problems with measuring and modeling musical ability has been accepted for publication and is forthcoming in *Musicae Scientiae* and work from the fifth chapter on using experimental methods as a means to understand melodic dictation is set to be published in a chapter for an upcoming *Routledge* book that has been tentatively titled *Understanding Melodic Dictation via Experimental Methods*. Once finished, the computational chapters are set to be submitted to the *International Society for Music Information Retrevial* and results from the experiments are set to be published as an article that includes both sets of experiments for a music perception journal. A review of melodic dictation literature as it relates to Karpinski's work as well as the computational model as it pertains to music theory pedagogy will also be published in a journal where the readership overlaps between music theory, music cognition, and music education. Current progress of the dissertation is available online where all the materials, software, text, and supplemental literature cited in the dissertation can also be accessed in able to facilitate more accessible research for anyone looking to become involved with this area of research.

A completed draft of the dissertation will be sent to the committee in early 2019, thus allotting two and a half months for revisions before the official submission date of March 18th. The dissertation will be defended in the weeks after submission at a time where all members of the committee are able to attend.

Select Bibliography

Berz, William L. 1995. "Working Memory in Music: A Theoretical Model." *Music Perception: An Interdisciplinary Journal* 12 (3): 353–64. https://doi.org/10.2307/40286188.

Butler, David. 1997. "Why the Gulf Between Music Perception Research and Aural Training?" *Bulletin of the Council for Research in Music Education*, no. 132.

Cowan, Nelson. 1988. "Evolving Conceptions of Memory Storage, Selective Attention, and Their Mutual Constraints Within the Human Information-Processing System." *Psychological Bulletin* 104 (2): 163–91.

———. 2010. "The Magical Mystery Four: How Is Working Memory Capacity Limited, and Why?" *Current Directions in Psychological Science* 19 (1): 51–57. https://doi.org/10.1177/0963721409359277.

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

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

Klonoski, Edward. 2000. "A Perceptual Learning Hierarchy: An Imperative for Aural Skills Pedagogy." *College Music Symposium* 4: 168–69.

Meyer, Leonard. 1956. *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

"National Association of Schools of Music Handbook." 2018. Reston, Virginia: National Association of Schools of Music.

Ortmann, Otto. 1933. "Some Tonal Determinants of Melodic Memory." *Journal of Educational Psychology* 24 (6): 454–67. https://doi.org/10.1037/h0075218.

Paney, Andrew S. 2016. "The Effect of Directing Attention on Melodic Dictation Testing." *Psychology of Music* 44 (1): 15–24. https://doi.org/10.1177/0305735614547409.

Pearce, Marcus 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): 378–95. https://doi.org/10.1111/nyas.13654.

Pembrook, Randall 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. https://doi.org/10.2307/3345259.

Taylor, Jack A., and Randall G. Pembrook. 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. https://doi.org/10.1037/h0094258.

Chapter 7

Computational Model

7.1 Levels of Abstraction

In his 2007 article *Models of Music Similarity* (Wiggins, 2007), 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 (Pearce, 2018).

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 (Saffran et al., 1999; Margulis, 2014). 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 (Huron, 2006).

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 (Karpinski, 2000, 1990) 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 (Pearce, 2005) and from cognitive psychology I draw from Nelson Cowan's Embedded Process model of working memory (Cowan, 1988, 2010) 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 (Lerdahl and Jackendoff, 1986; Narmour, 1990, 1992; Temperley, 2004) 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 (McDermott et al., 2016), other work suggests that experiencing pitches as discrete, categorical phenomena is categorized as a statistical human universal (Savage et al., 2015). 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 (Krumhansl, 2001; Levitin and Tirovolas, 2009).

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 Pearce

[1] As conceptualized as either a Schenkerian foreground (Schenker, 1935) or defined by Lerdahl and Jackendoff (1986)

(2018). 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 (Pearce, 2005, 2018). 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 (Shannon, 1948), 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 (Pearce, 2018), harmony– outperforming (Harrison and Pearce, 2018) sensory models of harmony (Bigand et al., 2014)–, and is also being developed to handle polyphonic materials (Sauve, 2017).

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 (Lerdahl and Jackendoff, 1986; Margulis, 2005; Narmour, 1990, 1992), 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 (Pearce, 2018), 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.

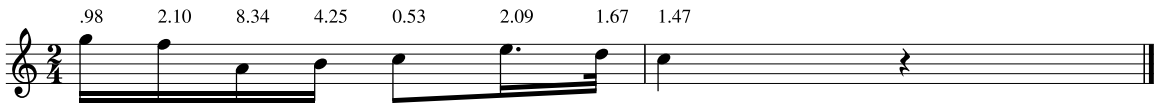

Figure 7.1: Cadential Excerpt from Schubert's Octet in F Major

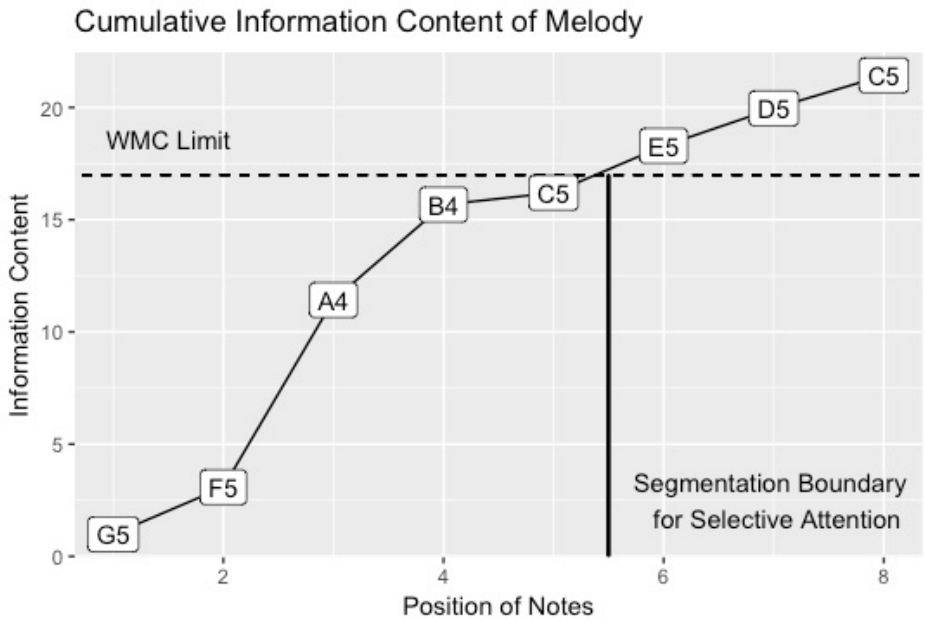

Figure 7.2: Cumulative Information in Schubert Octet Excerpt

In order to create this metric, I rely on IDyOM's use of information content (Shannon, 1948) 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 Pearce (2018) 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 (Schaffrath, 1995) 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 (Cowan, 1988, 2010). 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 (Cowan, 1988, 2010), 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 Baddeley and Hitch (1974) 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 (Karpinski, 2000, 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

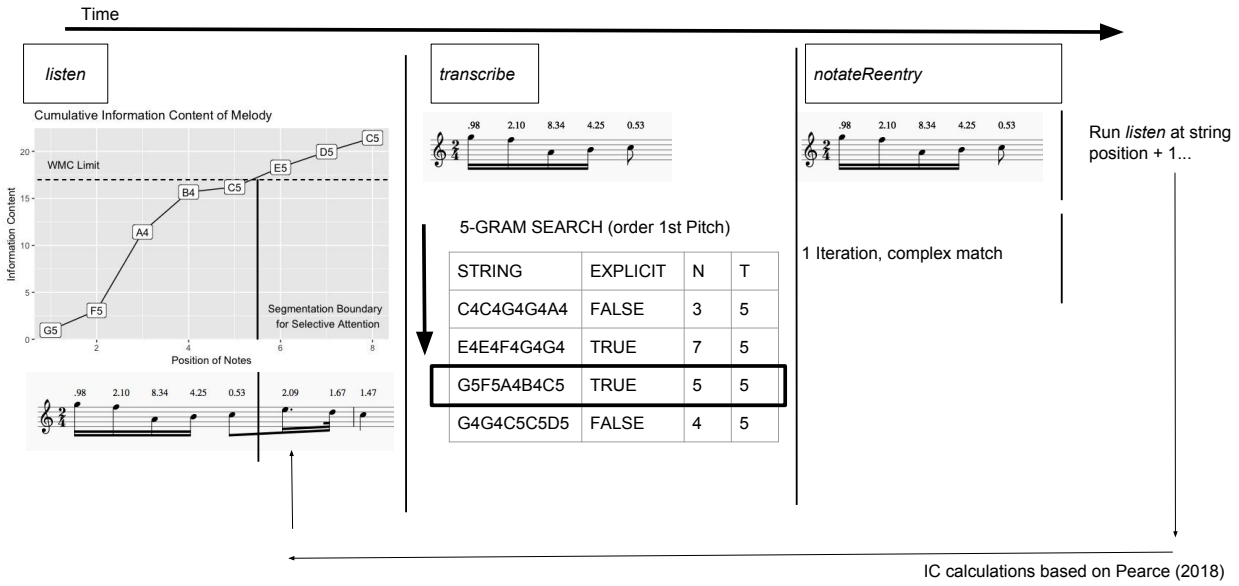

Figure 7.4: Model Example

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.