Modeling Melodic Dictation

David John Baker
2nd November, 2018
C4DM @ QMUL

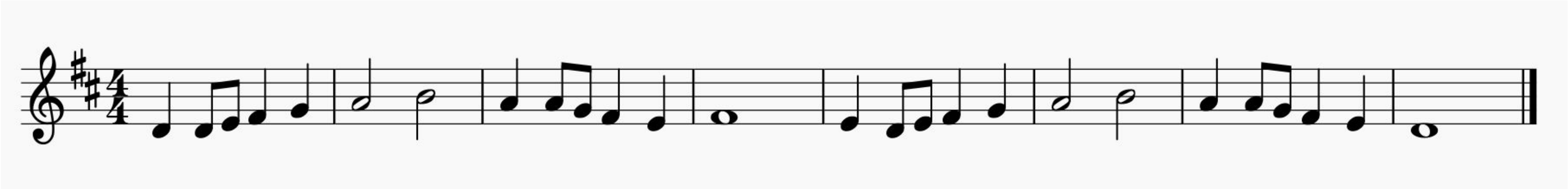
The Plan

I. Convince you music theory needs more research in melodic dictation
II. You can do this with **experimental** methods
III. You can do this with **computational** methods

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 it on paper in a limited time frame.

Example



Example

National Association of Schools of Music

HANDBOOK 2017-18

2. **Musicianship Skills and Analysis.** Students must acquire:

   a. An understanding of the common elements and organizational patterns of music and their interaction, the ability to employ this understanding in aural, verbal, and visual analyses, and the ability to take aural dictation.

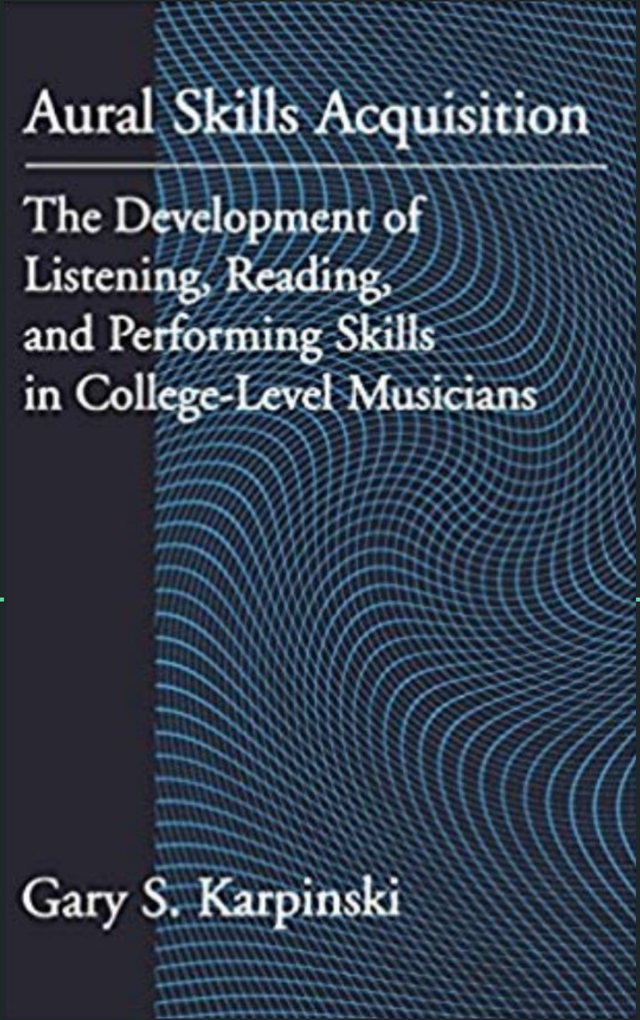
   b. Sufficient understanding of and capability with musical forms, processes, and structures to use this knowledge and skill in compositional, performance, analytical, scholarly, and pedagogical applications according to the requisites of their specializations.

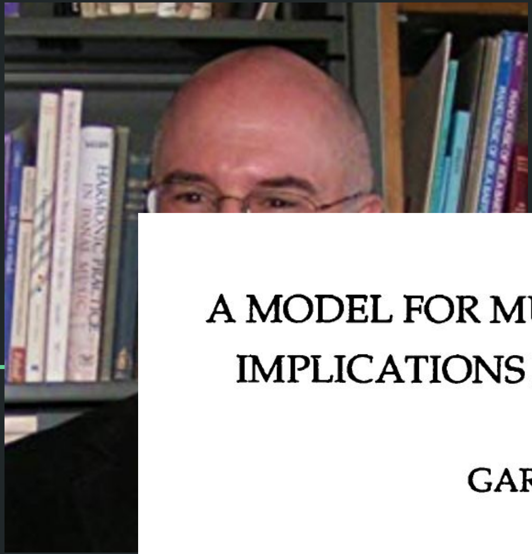
   c. The ability to place music in historical, cultural, and stylistic contexts.

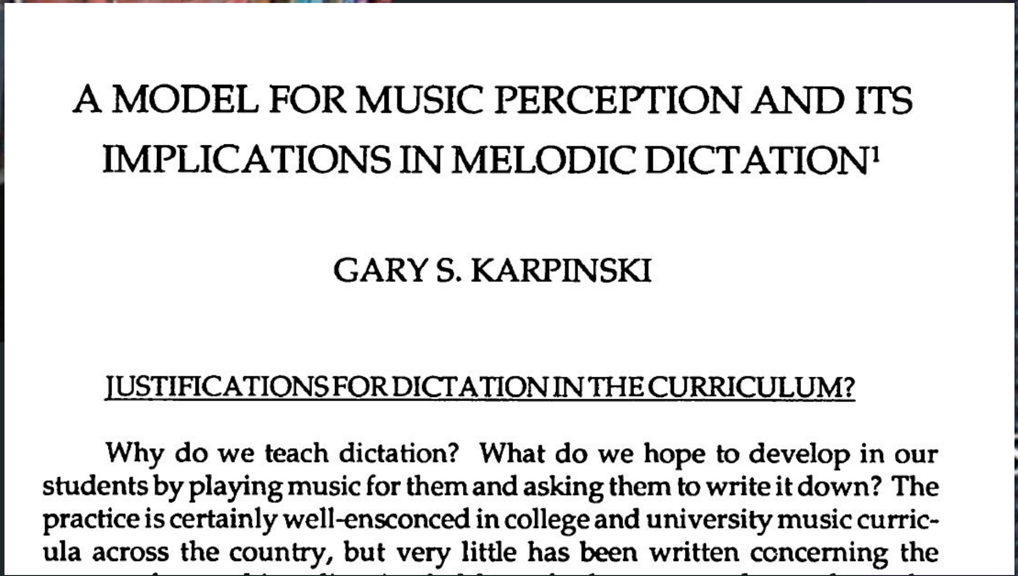
Dissertation Value?

Chapter 1

Significance of the Study


Aural Skills Acquisition
The Development of
Listening, Reading,
and Performing Skills
in College-Level Musicians
Gary S. Karpinski



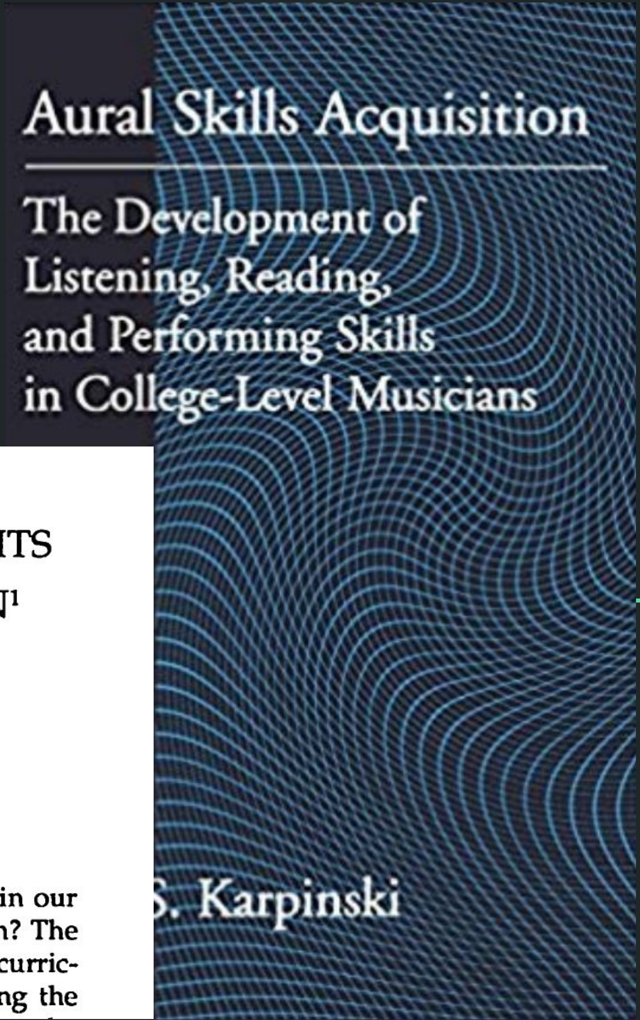


A MODEL FOR MUSIC PERCEPTION AND ITS IMPLICATIONS IN MELODIC DICTATION[1]

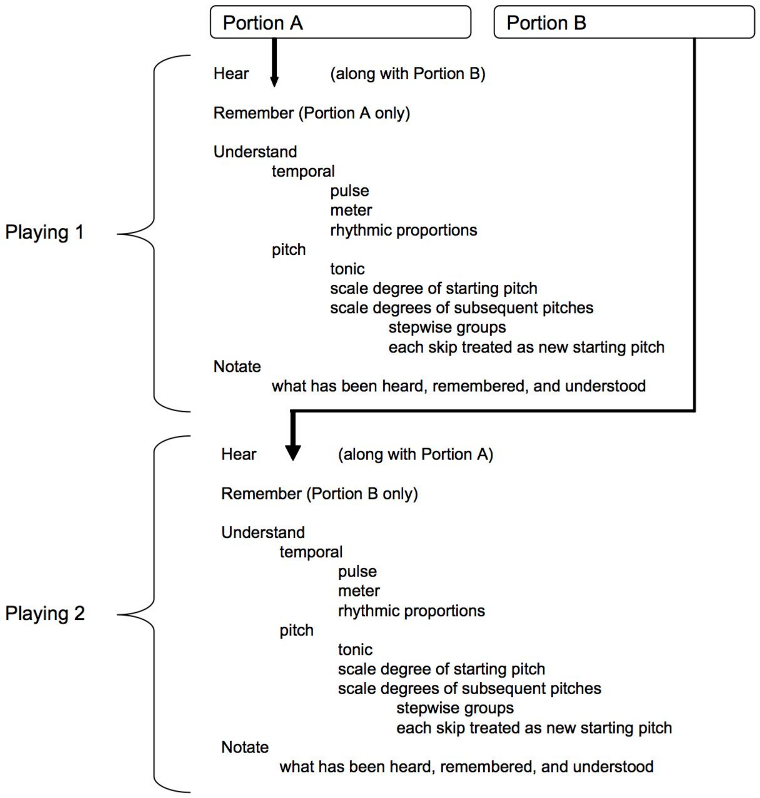
GARY S. KARPINSKI

JUSTIFICATIONS FOR DICTATION IN THE CURRICULUM?

Why do we teach dictation? What do we hope to develop in our students by playing music for them and asking them to write it down? The practice is certainly well-ensconced in college and university music curricula across the country, but very little has been written concerning the

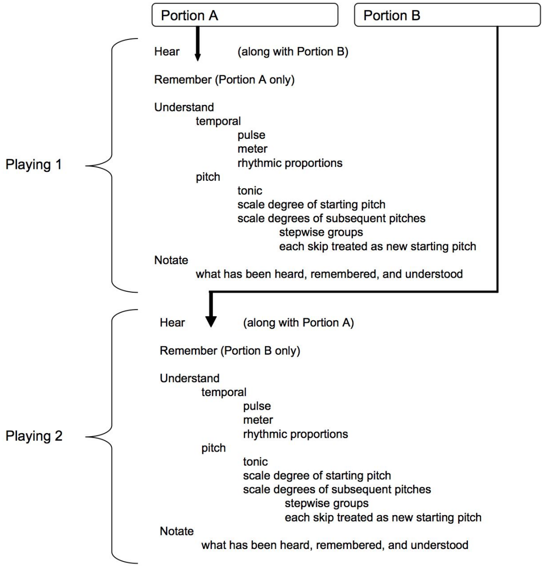


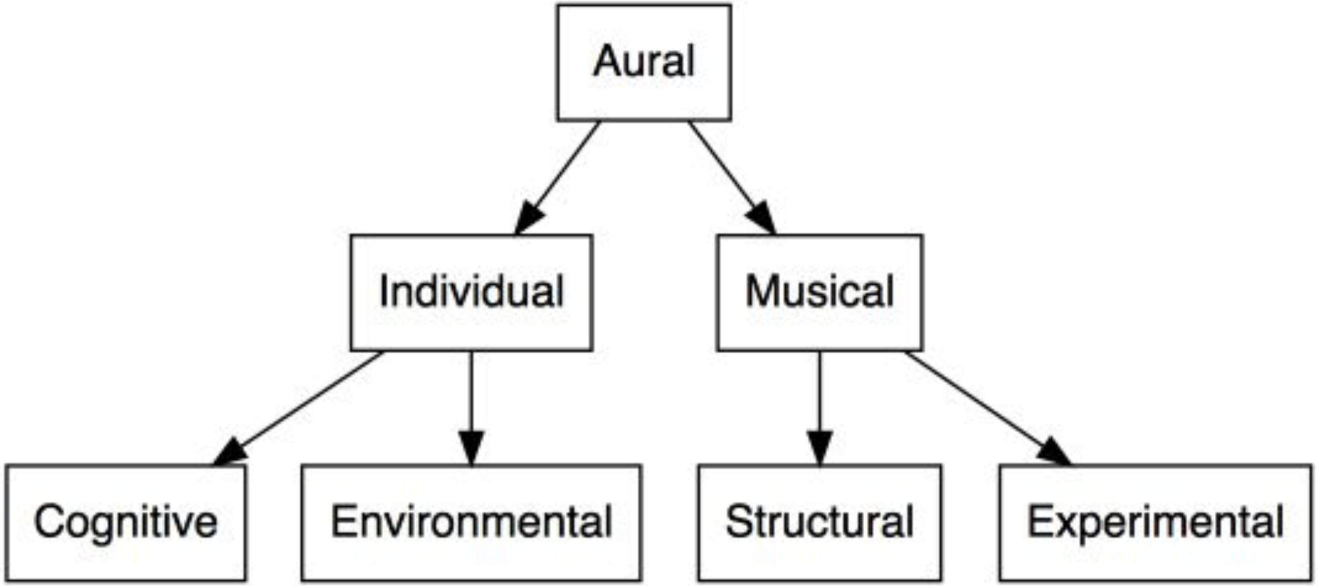
1. Hearing
2. Musical Memory
3. Understanding
4. Notation :||



Karpinski Model

- Fantastic for Pedagogical Applications
- Verbal Model
- *Idealized*
  - Agnostic to Individual + Musical Differences




Aural
Individual
Musical
Cognitive
Environmental
Structural
Experimental

Individual

Cognitive

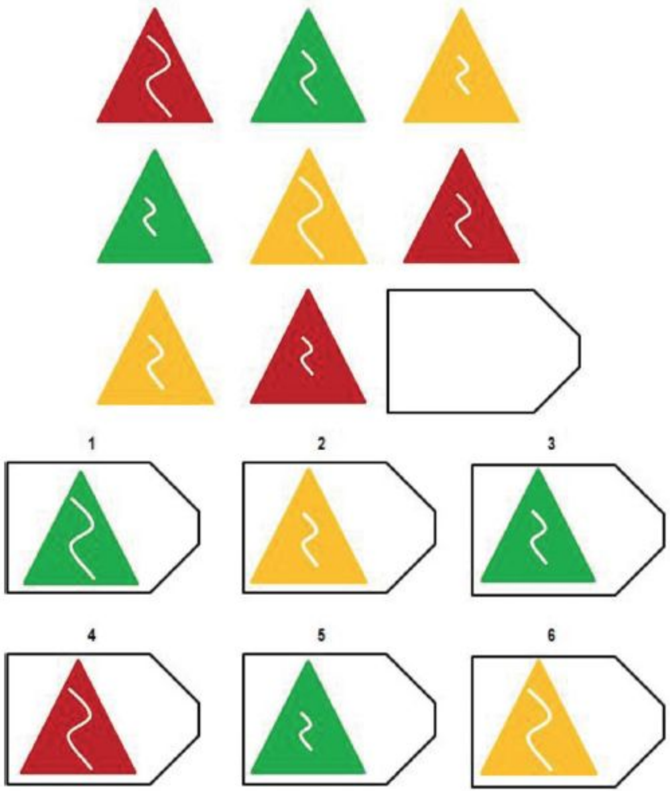
I. General Fluid Intelligence
II. Working Memory

I. General Fluid Intelligence (*Gf*)

Gf ➜ Ability to solve problems in novel situations, regardless of previous knowledge

Gc ➜ Ability to solve problems using acquired skills or knowledge

(Cattell, 1971 ; Horn, 1994)



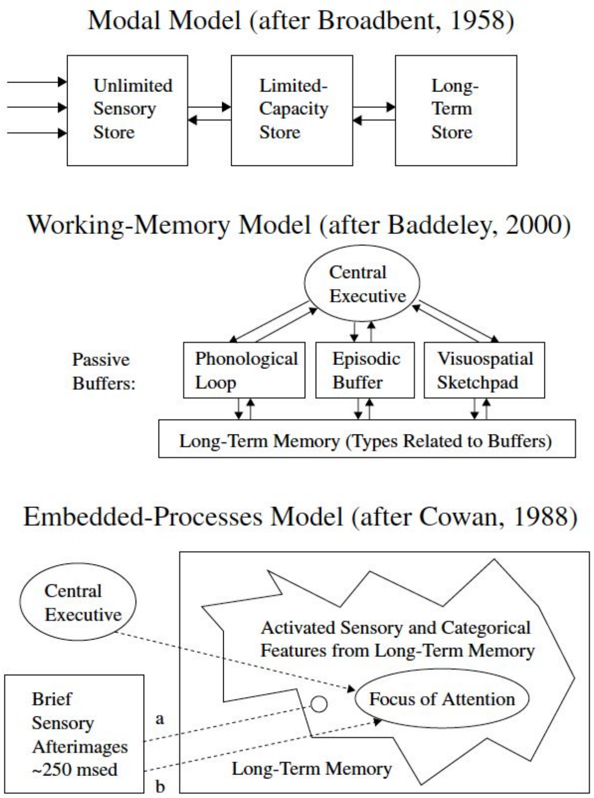
Matrix Reasoning (Kovacs and Conway, 2016)

Music and Intelligence

- Children and Adults with musical training score higher on measures of intelligence (Gibson, Folley and Park, 2009; Hille et al., 2011; Schellenberg, 2011; Schellenberg and Mankarious, 2012)
- 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)

II. Working Memory

“the relatively small amount of information that one can hold in mind, attend to, or, technically speaking, maintain in a rapidly accessible state at one time. The term working is meant to indicate that mental work requires the use of such information.” (Cowan, 2005)



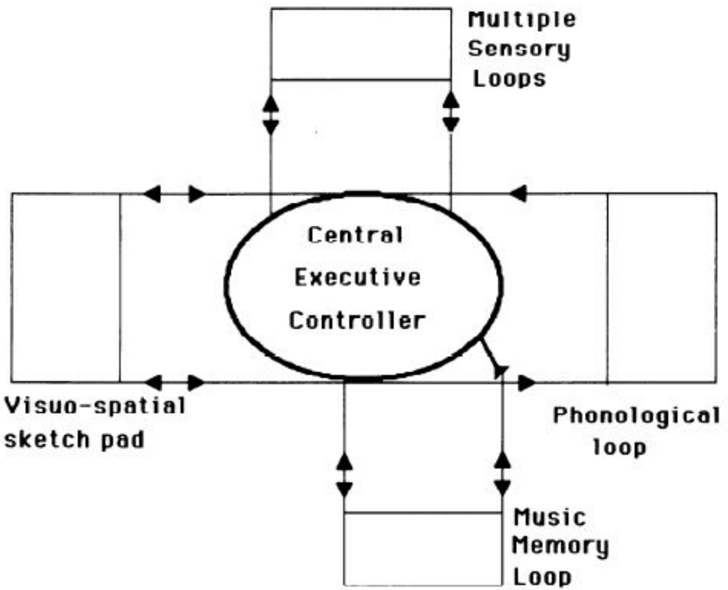


Fig. 1. Theoretical model of working memory based on the model of working memory by Baddeley (1990).

Lots of Models

Literature (Simple Span Tasks)

- Musicians generally perform better on tests of memory (Talamini et. al, 2017)
- WMC related to piano sight reading at lower performance levels
  - Kopiez & Lee 2006; Kopiez 2008
- WMC significant predictor of piano sight reading above and beyond practice
  - Meinz and Hambrick 2010
- Conductors greater flexibility although no difference in WMC
  - Wöllner and Halpern 2016
- Jazz musicians better at playing back, through negligible differences in WMC
  - Nichols, Wöllner, and Halpern
- WMC predictive in novel tapping task
  - Colley, Keller, Halpern 2018

Working Memory and Melodic Dictation

- Active manipulation of new, novel musical information?

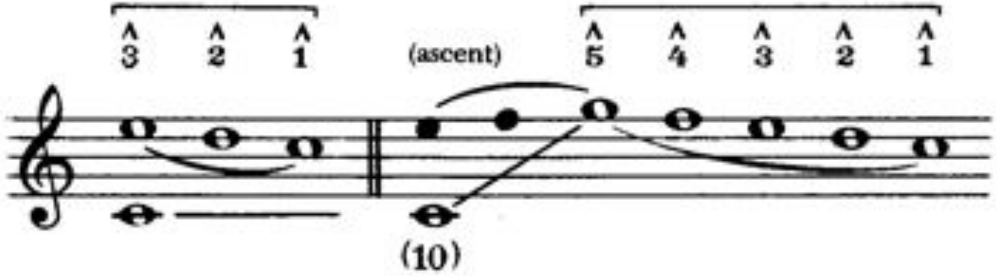
“Individual differences portrayed in some music aptitude tests may then represent not talent or musical intelligence but ability, reflecting differences in working memory capacity.” (Berz, 1995, p. 362)

Musical

Background:
fundamental structure

1

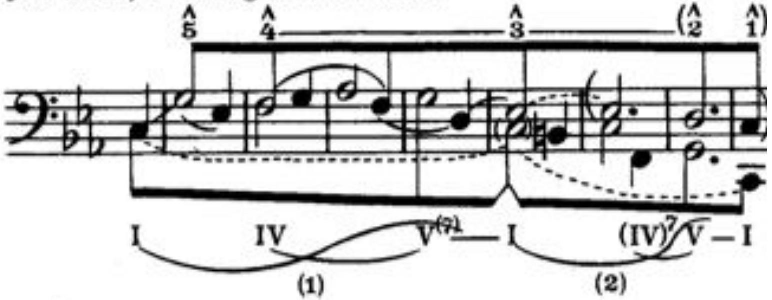
5

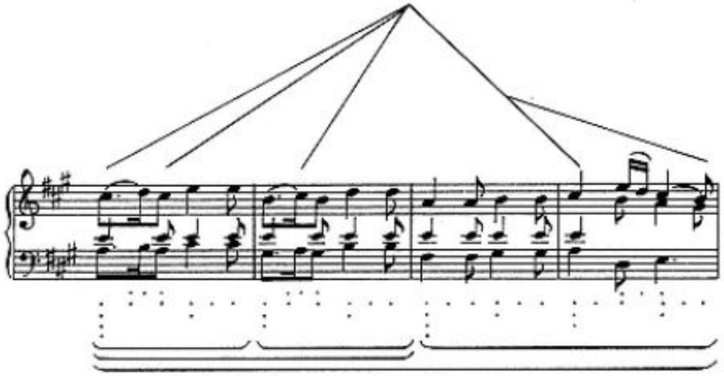


(See Fig. 19, *a.*)

1 J.S. Bach, Passacaglia in C Minor

20





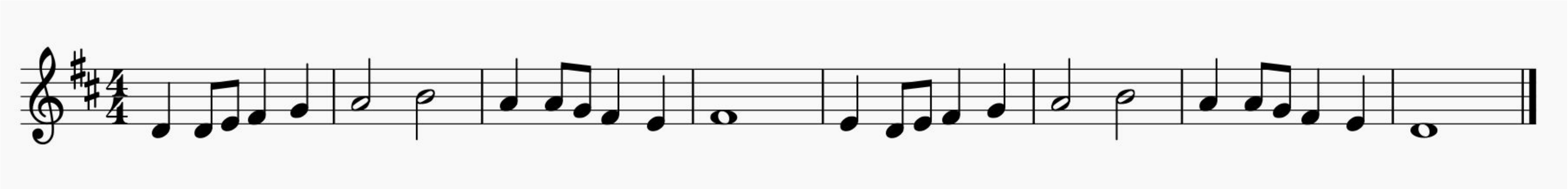
5.10

Abstracted Features

I. Static
II. Dynamic

Abstracted Features

I. Static (FANTASTIC)
II. Dynamic (IDyOM)



Mean Entropy	Mean Productivity	Mean Simpsons	p.range	i.entropy	note.dens	tonalness
.78	.41	.053	9	.44	1.85	.8539

FANTASTIC: Feature ANalysis Technology Accessing STatistics (In a Corpus): Technical Report v1.5

Daniel Müllensiefen

June 19, 2009

C (bars 23–24)

Position	1	2	3	4	5	6	7	8
Pitch	G_5	F_5	A_4	B_4	C_5	E_5	D_5	C_5
Probability	0.509	0.234	0.003	0.053	0.691	0.234	0.314	0.360
IC	0.98	2.10	8.34	4.25	0.53	2.09	1.67	1.47

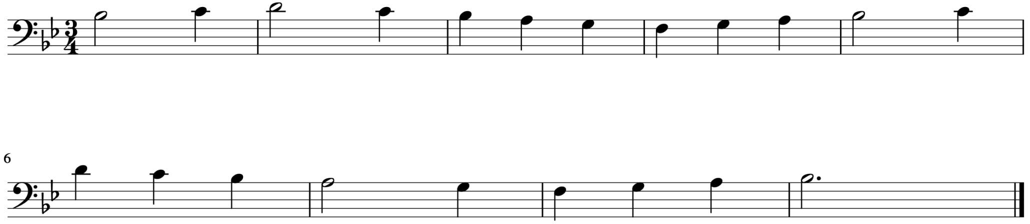
Dynamic (IDyOM)

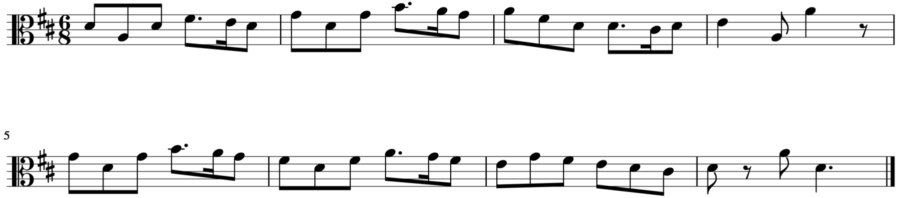
Experiment

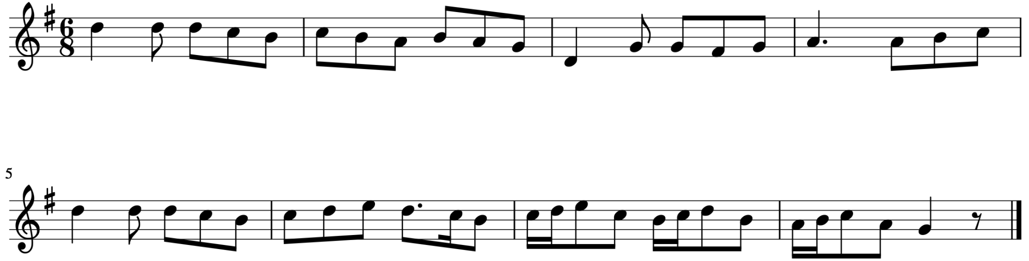
Hypotheses

- H1: Are all experimental melodies equally difficulty to dictate?
- H2: To what extent do the musical features of Note Density and Tonalness play a role in difficulty of dictation?
- H3: Do individual factors at the cognitive level play a role in the melodic dictation process above and beyond musical factors?


6


6


5


5

Methods I

- N = 43 → N = 39 participants
- Mean Age = 19.81, SD = 1.93, 15 Women
- Selected Four Melodies from corpus of 115 via FANTASTIC features

Melody	Note Density (ND)	Tonalness	Label
9	1.75	.71	Low ND, Low Tonal
34	1.66	.94	Low ND, High Tonal
95	3.91	.76	High ND, Low Tonal
112	3.73	.98	High ND, High Tonal

Procedure

- Melodic Dictation
  - 5x each, 20 seconds between hearings, 120 seconds after last hearing
- Aural Skills Questionnaire
- Bucknell Auditory Imagery Scale C (Halpern, 2015)
- Number Series (General Fluid Intelligence)
- Raven's Advanced Progressive Matrices (Gf)
- Rotation Span (Working Memory)
- Symmetry Span (WM)
- Goldsmiths Musical Sophistication Index
- STOMP/SES/Hearing

Lab Battery

+

A

$$1 + (2 + 1) = 0$$

TRUE OR FALSE

F

$$3 * (1 + 1) = 6$$

TRUE OR FALSE

J

$$1 + 2 = 4$$

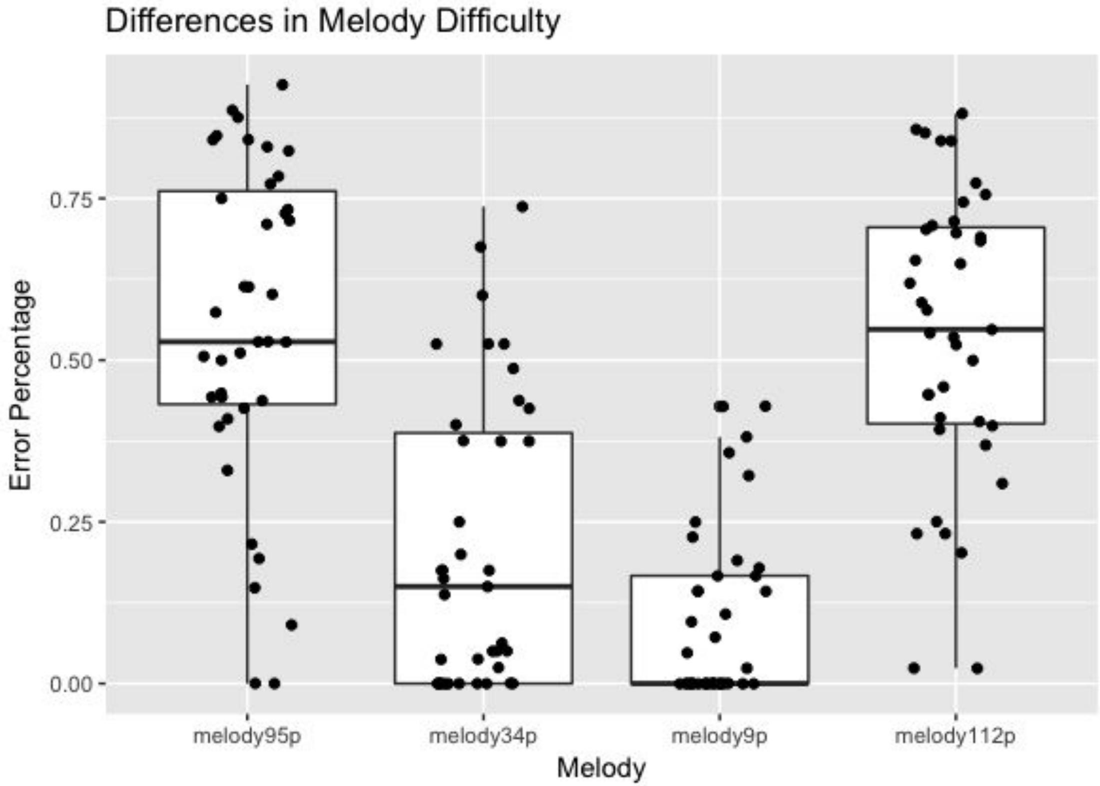
TRUE OR FALSE

+

A	E	B
T	F	R
J	Z	C

Select In Order Remembered!

Analysis -- RMANOVA



IV	P value	η2
Tonalness	.043	.01
Note Density	.001	.46
T X ND	.012	.02

H2 + H3

- General Linear Mixed Effects Models were run
  - No effect of WM
  - No Effect of Gf
- Did not continue with further analysis at time for inflated Type I Error

Future Experimental Work

- Run 2nd Experiment without 2x2 design
- Larger sample
- Remove 6/8 confound
- Use other FANTASTIC Features to re-analyze current data
- Still doesn't *explain* how this happens...

Computational Model

Musicae Scientiae
Discussion Forum 4A, 2007, 315-338

Models of musical similarity

GERAINT A. WIGGINS
Centre for Cognition, Computation and Culture
Goldsmiths College, University of London

Descriptive Theories → What happens

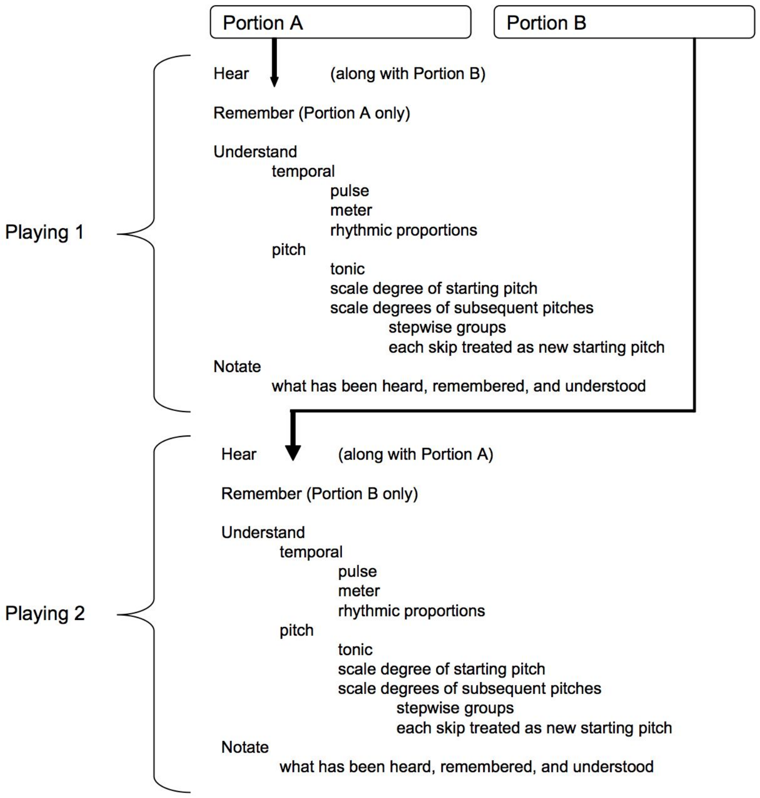
Explanatory Theories → What + How and Why!

1. Hearing

2. Musical Memory

3. Understanding

4. Notation :||



Computational Model

I. Prior Knowledge
II. Selective Attention
III. Transcription and Re-entry

Model Overview

- Prior Knowledge: previous knowledge an individual brings to the melodic dictation

- Selective Attention: Segments incoming musical information

- Transcription and Re-Entry: Pattern Matches and “Writes Down” melody

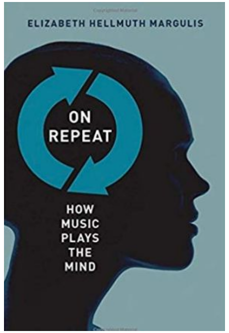
Theoretical Frameworks

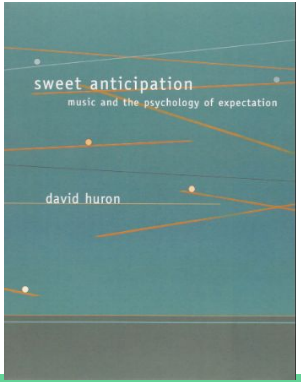
Theories

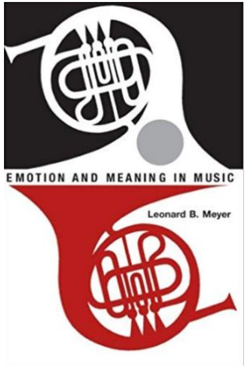
I. Statistical Learning ← → Pearce
II. Embedded Process ← → Cowan

Statistical Learning

- Emotion and Meaning in Music (Meyer, 1956)
- IDyOM -- Pearce (2005)
- Sweet Anticipation -- Huron (2006)
- On Repeat -- Margulis (2014)







Pearce, 2018

Statistical Learning Hypothesis:

“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

Probabilistic Prediction Hypothesis:

"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

Quantifying Unexpectedness

- IDyOM -- Computational model of auditory cognition based on SLH + PPH

C (bars 23–24)

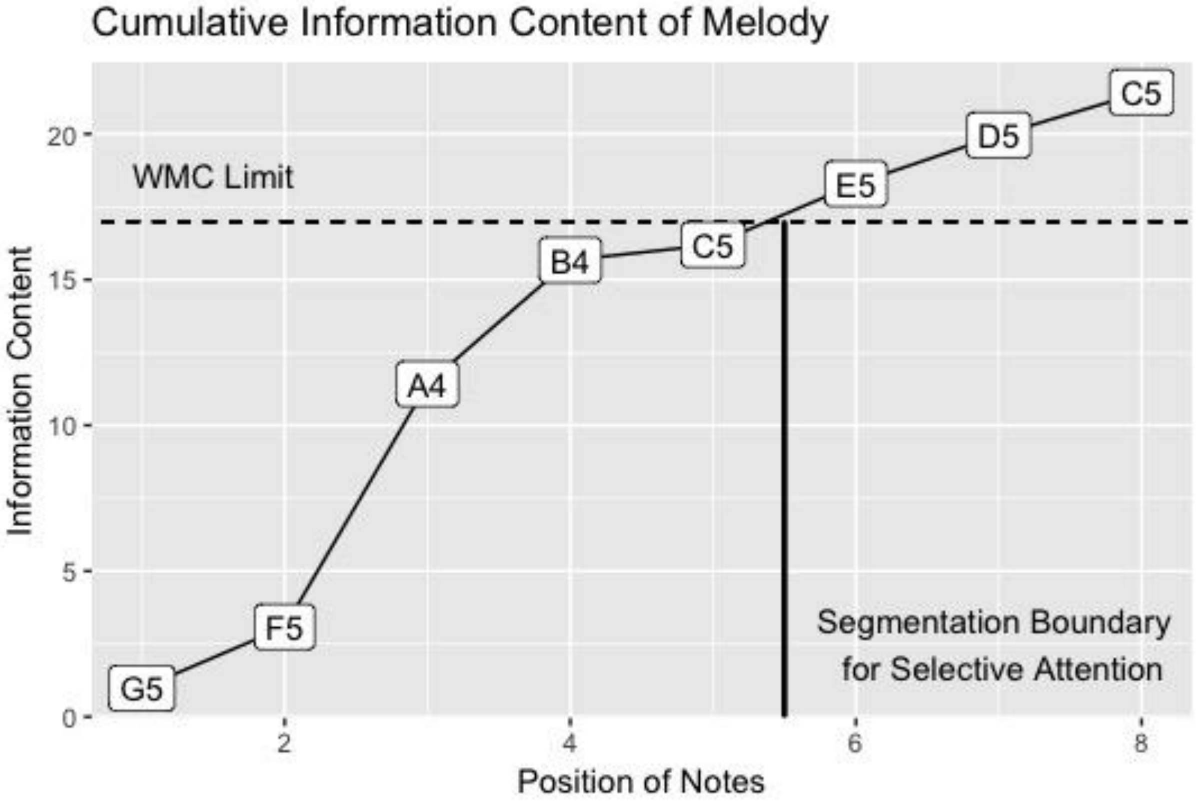
Position	1	2	3	4	5	6	7	8
Pitch	G_5	F_5	A_4	B_4	C_5	E_5	D_5	C_5
Probability	0.509	0.234	0.003	0.053	0.691	0.234	0.314	0.360
IC	0.98	2.10	8.34	4.25	0.53	2.09	1.67	1.47

Taken from Fig. 2 Pearce, 2018

C (bars 23–24)

Position	1	2	3	4	5	6	7	8
Pitch	G_5	F_5	A_4	B_4	C_5	E_5	D_5	C_5
Probability	0.509	0.234	0.003	0.053	0.691	0.234	0.314	0.360
IC	0.98	2.10	8.34	4.25	0.53	2.09	1.67	1.47

Information Content

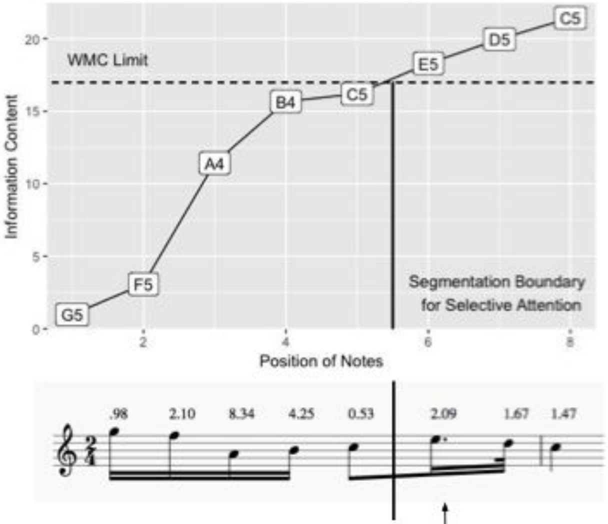


Information Content

Time

listen

Cumulative Information Content of Melody



transcribe

5-GRAM SEARCH (order 1st Pitch)

STRING	EXPLICIT	N	T
C4C4G4G4A4	FALSE	3	5
E4E4F4G4G4	TRUE	7	5
G5F5A4B4C5	TRUE	5	5
G4G4C5C5D5	FALSE	4	5

notateReentry

1 Iteration, complex match

Run *listen* at string position + 1...

IC calculations based on Pearce (2018)

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()

Predictions

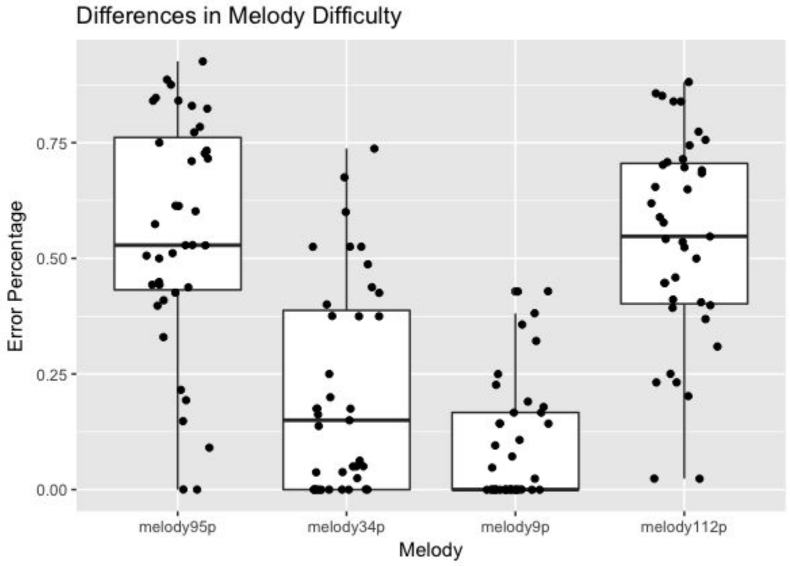
- Segments of melodies dictated relative to their distribution in Prior Knowledge
- Higher WM span people do better on dictation
- Interval level dictation will result in more consistent, but less effective dictations
- Difficulty of melody can be predicted by frequency of n-gram distributions
- Atonal melodies made of more frequent n-grams should be easier to dictate than tonal melodies with less frequent n-gram distributions
- Higher exposure to sight singing results in more learned patterns, thus ability to dictate larger patterns of music

Limitations

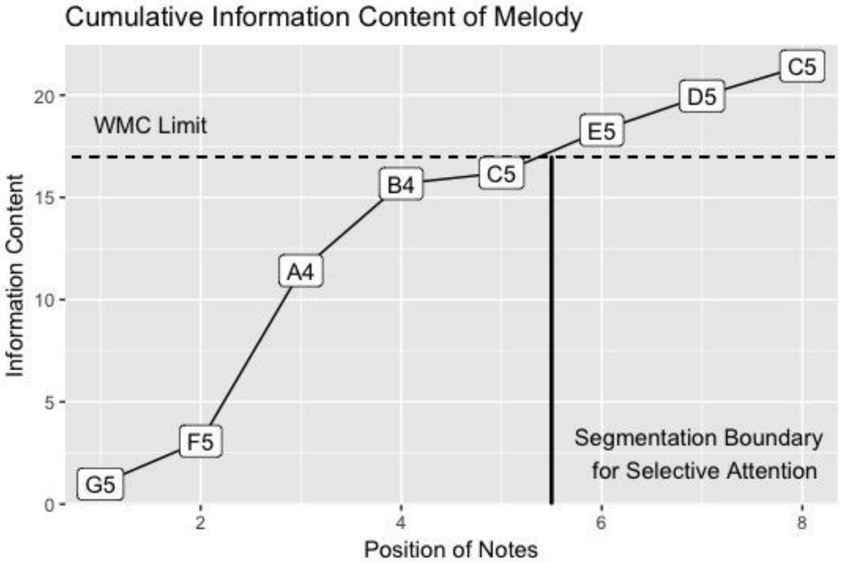
- How to determine the “Prior Knowledge”
  - Set based on melodies learned per semester?
  - Could look at data, reason backwards? Tedious?
  - What counts as an explicitly learned interval?
- Question of scoring and grading
  - Score same to human data
- Tonic Inference
- Determining IC Thresholds
  - Gating experiments?
  - Isolating musical features?

Takeaways

.98
2.10
8.34
4.25
0.53
2.09
1.67
1.47



Differences in Melody Difficulty
Error Percentage
0.00
0.25
0.50
0.75
melody95p
melody34p
melody9p
melody112p
Melody



Cumulative Information Content of Melody
Information Content
0
5
10
15
20
WMC Limit
G5
F5
A4
B4
C5
E5
D5
C5
Segmentation Boundary
for Selective Attention
2
4
6
8
Position of Notes

Thank Yous