

1 Computational Tools for Aural Skills Pedagogy

2 David John Baker¹

3 ¹ Louisiana State University

4 Author Note

5 David John Baker completed this work while a Ph.D. student at Louisiana State
6 University. He currently works for Flatiron School in London, England.

7 Maybe something else can be added here.

8 Correspondence concerning this article should be addressed to David John Baker, 131
9 Finsbury Pavement. E-mail: davidjohnbaker1@gmail.com

Abstract

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular study.

One sentence summarizing the main result (with the words “**here we show**” or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

One or two sentences to put the results into a more **general context**.

Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

Keywords: aural skills, computational musicology, survey methods, assessment

Word count: X

Computational Tools for Aural Skills Pedagogy

Introduction

Which of the melodies in Figure 1 would be more difficult for a student to dictate?

Phrase level Within those phrases, appears that there is smooth melodic contour.

Melody is scalar, lacks any syncopation, and non-diatonic notes seem to just be ornamental embellishments that happen at predictable place in antecedent phrase of parallel period. This phrases structure leads itself to basic harmonization for undergraduates.

In contrast, melody B lacks the qualities mentioned above. It does not have mirrored phrase structure. There are moments of scalar movement, but often after leaps. Syncopation makes for unstable. And introduction of the non-diatonic tone here probably would give the listener the impression that the F# might actually be diatonic as there has not been any time for a tonal center to be established.

I would find it probably fair to wager that most aural skills pedagogy would bet that students would be better at dictation melody A rather than melody B. And in discussions among aural skills instructor practitioners, the above analysis is probably good enough in terms of providing reasons to answer the question of if one melody is just harder to dictate than the other.

But what if we wanted to know not just a matter of rank/less than greater than? What if the question became a matter of degree? Imagine a scenario where need to start to get idea for standardization. Say more here why. How much more difficult is melody B rather than melody B? Possible to take an empirical stance on this and say if you really wanted an answer you could have students dictate both, score them, then make some sort of ratio comparison. Problem with this is that this answer would be dependent on the parameters of the dictation. How good are the students? And would this result generalize if done multiple times with both students in their first few dictations compared to those

completing their university music training? If we were to standardize, this problem exemplifies a common problem that aural skills practitioners deal with. It's easy to make relative rank order assessments in our classroom and we rely on our experience and intuition in order to do this. And while this works on a per institution, even per classroom basis, lack of standardisation makes it very difficult to communicate across other educators about agreeing on what we can ask students to do and be fair in our assessments since aural skills is so personalized. Without an objective standard, we lack the language as community for common denominator to talk about the degree to which we can expect of students. So what is a girl to do?

Historically, thing to do here is to cross the bridge to music cognition to help inform pedagogical practices. Thesis here will be to show how tools from computational musicology can help to standardise this problem and the benefits of doing so.

Objectiveness

One of most important things we can do as teachers is make sure that we are teaching students material they don't know and then after than checking for understanding that they learn the materials. When we check for understanding, via either formative or summative assessments, its important to make sure that what we are testing the student on was within their control. This ensures a sense of fairness to the student and could add sentence here about how this is important for sense of student morale in the classroom (teaching tech).

As shown from example above, coming to some sort of objective understanding of what can be said for music is very difficult. What is considered basic will differ from institution to institution. Also whatever standards are set, most people report that students entering music schools are unprepared for whatever they are about to do (Cite). Of course the answer here is not to try to assign universal scoring rubric to what students

77 should or should not be able to do. What is thought of as a noivce really does not mean
78 much given the diversity of insitutions and students that teachers of aural skills will all
79 experience.

80 As it should be, in the classroom setting, what the level of difficulty is for the
81 classroom is set by the instructor given their expertise. This depends on the aims and
82 learning goals of each lesson and the higher level of the course. This is OK for doing
83 locally, last thing we need is universal standards in which we remove ability to do custom
84 or bespoke lesson planning. The problem is that as a field, without any sort of objective
85 standard, it becomes very hard to communicate common things and preseve institutional
86 knowledge. For this to happen, you need some sort of common denominator to act as
87 difficulty rosetta stone.

88 As stated above, this typically comes from crossing the bridge to music cognition.
89 There is a history of this where this is studied. Take a look at Ortmann, taylor pembrook.
90 Plus all the stuff with the educaiton people. And finally there has been convergance on this
91 from computational musicology with features. A feature is way to summarize the contents
92 of melody after it has been digitized into discrete tokens, bascially when you make it a
93 bunch of notes. Features that readers from music theory and education might be familiar
94 with are range or the more abstract idea of global key for melody. While something like
95 range would be objective, can see that something like key gives room for debate. Much of
96 work on features is inspired by computational linguistics. For example, also might be
97 familiar with nPVI. Cite paper saying its misused. But importantly is that it gives
98 objective measure.

99 As discussed in Baker, can either be static or dynamic. Static features are just
100 summary of the whole melody. Advantage of this is that have single number to describe.
101 For example, range is xyz. Also sometthing like interval entropy, which here is layman
102 term, and is predictive of xy'. These features have been shown to predict in THESE

EXPERIMENTS. This gives objective measure.

Negative sides of this is that does not match with phenomenological experience. Assumes that melody is almost heard in suspended animation. Make joke about key here. Note also that here get the problem where if suspended animation, then can have situation arise in Figure 2 where two melodies can have exact same rating on some metrics that assume summary but clearly at not represented at phenomenological level. And also some of the features like note density are going to interact with tempo. And problematically, as noted in previous literature, all features going to correlate. Hard to change one without others. There are some ways around this (lasso, ridge) for prediction, or could take PCA approach.

On other side of this, can also look at dynamic models. Here each token that was summarized gets its own value associated with it. Clearest example of this is work from Marcus Pearce and IDyOM model. Def of IDyOM here. Essentially is able to make some sort of surprise rating based on information content. Regardless of buying metaphor of brain as computer and discussion around that, variable can be predictive without invoking any sort of casual or mechanistic claims about cognition. For example, show what IDyOM would have put for melody A and B in Figure 2 along with table of select features here. Advantages of this lead to things like FFH in DJB. Disadvantages of this is need larger ML model to get it to run and get values and assume some sort of stochastic, not deterministic like FANTASTIC.

Regardless of static versus dynamic, the idea here is that the features will provide some sort of grounding since they computed with paper trail. More importantly, can now answer questions of degree that were the problem before.

So then this leads to the question, now that we have common language, does it actually map onto what intuitive experience. These would not be helpful if they were not predictive of expert intuition. Next thing to do here would be to see if these measures will

map mathmatically onto generalised cases like comparing the difficulty of dictation between melodies in Figure 1. Before going on, need to also note that have been saying difficulty and complexity as same thing. Whereas complexity in this context, i define to mean just the objective number from the computational measure, difficulty is going to relate but not be complexity. Difficulty here is going to be more empirical resulting from student performance. And note that the difficulty is emergent empirical feature that will interact with how the melody is performed. For example, any less complex melody would become more difficult if played at either very fast or very slow tempi.

So how is it possible to asses if one is predictive of the other and this actually would be helpful? Next I present a survey ran to see if there were some sort of relationships here.

Methods

- Methods go here for this
- First look at how features map on global level to just trajectory of book
- Not only that, but if you ask people to do the thing I am claiming that it does, agreement
- Here is all that data

Discussion

- Have shown here that tools from computational musicology provide good model for agreement
- Of course not perfect and should not be imposed as global standards
- But tacking down this one part will allow more context when start to talk about difficulty in the actual experience
- Benefits here are that we can begin to understand the meachanisms of this
- If we know it better, become better teachers
- And then point of all of this is if we know btter, we help students in the long run.

154 **Data analysis**

155 We used R (Version 3.6.2; R Core Team, 2019) and the R-package *papaja* (Version
156 0.1.0.9942; Aust & Barth, 2020) for all our analyses.

References

157

158 Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*.

159 Retrieved from <https://github.com/crsh/papaja>

160 R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna,

161 Austria: R Foundation for Statistical Computing. Retrieved from

162 <https://www.R-project.org/>