Spare Writing

David John Baker

11/04/2020

## Modeling Complexity

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

Attempting to use features of a melody to to predict how well a melody is remembered has a long history. In 1933, Ortmann put forward a set of melodic determinants that he asserted predicted how well a melody was remembered. These features such as a melody’s repetition, pitch-direction, contour (conjunct-disjunct motion), degree, order, and implied harmony (chord structure) were deemed to affect the melody’s ability to be remembered [@ortmannTonalDeterminantsMelodic1933].

Since Ortmann, pedagogues such as Taylor and Pembrook have expanded on this research, finding significant effects of musical features such as length, tonality, as well as type of motion as well as an effect of experimental condition [@taylorStrategiesMemoryShort1983]. Following up on Taylor’s investigation, @pembrookInterferenceTranscriptionProcess1986 found evidence corroborating Ortmann’s initial claims that his four major determinants (repetition, note direction, conjunct-disjunct motion, degree of disjunctivness) had a significant main effects on an individual’s ability to take dictation, yet note that these values do not exhaustively explain the findings. In their discussion, they also note the problems of completely isolating the effects of certain musical features as when you change one parameter, others are also subject to change. When looking at changes in structural elements of melodies, there is a collinearity issue among features. Not only does this problem exist within features of melodies, but also among participants. In reflecting on other factors that might contribute to their results, the authors note

Clearly, a complete hierarchy of determinants would constitute a very long list, because not only would the many melodic structures be included, but also their interactions with subject and environmental variables. The ones included in the present study (musical experience, melodic carryover, and response method) provided evidence that the melodic determinants are not constant; rather, they vary as a function of the subject and environmental factors, which in turn can have significant effects on music discrimination and memory. (p. 33) The authors later in the article go on to stress that future work should both replicate their findings as well as expand their modeling parameters. They call for a larger sample, a broader spectrum of musical experiences, and to investigate more musical features.

Since then some researchers have employed using features of the melodies to predict a behavioral measure in experimental settings. Not using as extensive of a battery as Ortmann, Taylor, or Pembrook, researchers in music psychology such as as @akiva-kabiriMemoryTonalPitches2009, @dewittRecognitionNovelMelodies1986, @eerolaExpectancySamiYoiks2009, @schulzeWorkingMemorySpeech2012 have used the number of notes in a melody as a successful predictor of difficulty in melodic perception and discrimination tasks. Expanding on just using frequency of note counts, @harrisonModellingMelodicDiscrimination2016 instead of looking at single measures of melodic complexity, addressed the melodic collinearity issue noted by Taylor and Pembrook by using data reductive techniques to derive a single complexity measure found to be predictive in their statistical modeling deriving these measures from the FANASTIC toolbox [@mullensiefenFantasticFeatureANalysis2009]. Following this research, @bakerPerceptionLeitmotivesRichard2017 also incorporated a similar measure of complexity in their model of leitmotiv recognition in which they predicted recall rates in a recognition paradigm.

Each of these examples operationalizes some feature of the melody with a quantitative, numerical proxy that is assumed to be able to be mapped to perception. Ortmann referred to these as determinants, while others such as Müllensiefen refer to them as features [@mullensiefenFantasticFeatureANalysis2009]. Since the word feature refers to a ‘distinctive attribute’, I will use this terminology throughout the rest of the chapter, though note that other terms have been used.

### What Are Features?

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

The nPVI began as a measure of rhythmic variability in language [@grabeDurationalVariabilitySpeech2002]. Shown below, the nPVI quantifies the amount of durational variability in language. It works by comparing the variability of vowel length compared to syllable length

where is the number of vowels in an utterance and is th duration of the item and has been used in musical contexts [@vanhandelRoleMeterCompositional2010].

In linguistics, the nPVI has been used to delineate quantitative differences between stress and syllable timed languages. Recently in the past decade, music science researchers have used the nPVI to attempt to investigate claims about the relationship between speech and music [@danieleInterplayLinguisitcHistorical2004; @patelStressTimedVsSyllableTimed2003; @vanhandelRoleMeterCompositional2010]. While results are mixed regarding the nPVI’s predictive ability and there have been recent calls to limit the measure’s use [@condit-schultzDeconstructingNPVIMethodological2019], it does serve as a very good example of a computational derived measure. Just like summarizing the range of a melody by subtracting the distance between the lowest and highest notes, the nPVI summarizes a phrase and importantly assumes that this measure is representative of the entire phrase the calculation was performed upon.

In computational musicology, features of melodies can generally be classified into two main types: static and dynamic features. Static features compute a summary measure over the entire melody while dynamic features calculate values for each event onset in a melody. One of the most complete set of static computational measures as applied to music perception come from Daniel Müllensiefen’s’ Feature ANalysis Technology Accessing STatistics (In a Corpus) or FANTASTIC toolbox [@mullensiefenFantasticFeatureANalysis2009].

FANTASTIC is a program…that analyzes melodies by computing features. The aim is to characterise a melody or a melodic phrase by a set of numerical or categorical values reflecting different aspects of musical structure. This feature representation of melodies can then be applied in Music Information Retrieval algorithms or computational models of melody cognition. (p. 4) Drawing from fields both central and peripheral to music science, FANTASTIC computes a collection of 38 features to analyze features of melodies and joined a large and continuing tradition of analyzing music computationally [@lomaxCantometricsApproachAnthropology1977 ;@eerolaExpectancySamiYoiks2009; @huronHumdrumToolkitReference1994; @lartillotMatlabToolboxMusical2007; @lomaxCantometricsApproachAnthropology1977; @mcfeeLibrosaAudioMusic2015; @steinbeckStrukturUndAhnlichkeit1982]. Additionally, FANTASTIC also provides a framework for comparing the features of a melody with a parent corpus from which the melody is assumed to belong similar to a sample-population relationship.

### Back to the Classroom

Returning to the Aural Skills classroom, many of these features can be used to approximate the previously established intuitions of complexity as agreed upon by theorists. Below in Figure @ref(fig:corfeature), I plot the the mean difficulty and grammar ratings given by experts for each melody in the experimental sample against each of the output of FANTASTIC’s features by correlating the two measures. Additionally, Table @ref(tab:corfeaturetable) displays the five strongest positive and negatively correlated features of FANTASTIC’s output with the ground truth, expert ratings.

# {r corfeature, echo=FALSE, fig.cap="FANTASTIC Computations Correlated with Expert Ratings",fig.align='center', out.width="100%"} # knitr::include\_graphics("img/FantasticExpertPlot.png") #

Strong Features

|  |  |  |
| --- | --- | --- |
| Feature | Difficulty | Grammar |
| i.abs.std | 0.89 | -0.82 |
| i.abs.mean | 0.87 | -0.91 |
| setp.cont.loc.var | 0.97 | -0.74 |
| i.entropy | 0.85 | -0.74 |
| p.entropy | 0.84 | -0.72 |
| d.median | -0.19 | 0.22 |
| d.eq.trans | -0.20 | 0.04 |
| mean.Yules.K | -0.43 | 0.40 |
| tonalness | -0.48 | 0.44 |
| mean.Simpsons.D | -0.57 | 0.50 |

From Figure REF and Table REF, there are some features that share a strong relationship with the ground truth of the expert intuitions. The top five features that correlate most strongly with the expert ground truths are related to the intervallic content of a melody. The first two features, i.abs.std and i.abs.mean are derived measures using absolute interval distance computations. The other top three features, step.cont.loc.var, i.entropy, and p.entroy are related to entropy measures. Of the negatively correlated features, two linguistically derived measures mean.Yules.K and mean.Simpsons.D both correlate with perceived difficulty, as does a measure of tonalness which in FANTASTIC is based on the Krumhansl key profiles [@krumhanslCognitiveFoundationsMusical2001].

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

# {r featurecorrelations, echo=FALSE, fig.cap="Visualization of the Problem of Melodic Feature Collinearity",fig.align='center', out.width="100%"} # knitr::include\_graphics("img/FANTASTIC\_collin.png") #

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

In REF we see computational evidence of claims made by @taylorStrategiesMemoryShort1983 when reviewing exactly what features might contribute to the degree of difficulty from a melodic dictation. Given this collinearlity problem, it becomes very difficult to be able to isolate the effect of one feature of the melody. One way to begin to understand these relationships would be to be to build statistical models that are able to partition covariance structures such as the general linear model when used in the context of multiple regression. Another method, as mentioned above, could instead take a more exhaustive, but less explanatory approach forward and follow past research [@bakerPerceptionLeitmotivesRichard2017; @harrisonModellingMelodicDiscrimination2016], that uses data reductive techniques such as principal components analysis to obtain more accurate predictive measures of complexity.

# {r univariatecow, echo=FALSE, fig.cap="Univariate Feature Models",fig.align='center', out.width="100%"} # knitr::include\_graphics("img/univariate\_cow.png") #

In Figure REF , I plot eight features extracted via the FANTASTIC Toolbox. The figure plots linear models of each feature compared against the expert ratings of difficulty. I additionally list the Pearson correlation coefficient for each model. From the plot, it is evident that some features correlate much stronger with the ground truth features than others. For example, pitch.entropy correlates with the ground truth data . Not only that, but the model is not being driven completely by outliers. While some points fall below the regression line, extreme values are not driving this effect. A similarly strong relationship is evident with the step.cont.local.var variable. In line with work by Dowling, this provides further evidence that contour changes have a significant impact on how people hear melodies [@dowlingScaleContourTwo1978]. In exploring these relationships in multivariate context, when I combined the top four variables from REF in a linear multiple regression model, the model was able to predict a high degree of variance . While this model is explicitly exploratory, this dataset will serve as a foundation to build future theories to test.

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

Although this approach has been relatively successful at modeling expert ratings, using FANTASTIC’s various linear combinations of these features does have important limitations. One of the most obvious limitations is that FANTASTIC’s measures tacitly assume listeners recall melodies in some sort of perceptual suspended animation. Illuminating this problem using a more tangible example, again returning to melodies FIGURE 1 and FIGURE 1, when the full set of FANTASTIC features are computed on both, the both melodies are computational equivalent in their range, pitch entropy, durtational range, durational entropy, length, tonalness, tonal clarity, tonal spike and stepwise contour global variation. This computation arises from computing a summary measure over the melody and not modeling it in terms of real time perception. In order to have a more phenomenologically appropriate model that incorporates computationally derived features, it is important to also consider dynamic models of music perception when modeling difficulty. Following up on another finding from this section, it also is worthy of mention that the variables with the strongest predictive powers tend to be those associated with information content. In the next section, I explore how using a dynamic approach such as Marcus Pearce’s implementation [@pearceConstructionEvaluationStatistical2005; @pearceStatisticalLearningProbabilistic2018a] of a multiple viewpoints model [@conklinMultipleViewpointSystems1995], might provide more insights into understanding the aural skills classroom.

### Dynamic

The Information Dynamic of Music (IDyOM) model of Marcus Pearce is a computational model of auditory cognition [@pearceStatisticalLearningProbabilistic2018a]. IDyOM is based on the assumption put forward by Leonard Meyer that musical style can be understood as a complex network of probabilistic relationships that underlie a musical style and is implicitly understood by a musical community [@pearceAuditoryExpectationInformation2012; @pearceConstructionEvaluationStatistical2005; @pearceStatisticalLearningProbabilistic2018a] that incorporates a multiple viewpoint framework [@conklinMultipleViewpointSystems1995]. Unlike measures from FANTASTIC, which calculate summary statistics based on melodic features, IDyOM works by calculating measures of expectancy of an event based on a predefined set of musical parameters that the model was trained on. As mentioned in Chapter 2, the IDyOM model relies on two important theoretical assumptions based on two neural mechanisms involved in musical enculturation: the statistical learning hypothesis and probabilistic prediction hypothesis. According to Pearce, the Statistical Learning Hypothesis (SLH) states that:

musical enculturation is a process of implicit statistical learning in which listeners progressively acquire internal models of the statistical and structural regularities present in the musical styles to which they are exposed, over short (e.g., an individual piece of music) and long time scales (e.g., an entire lifetime of listening). (Pearce, 2018) The logic here is that the more an individual is exposed to a musical style, the more they will implicitly understand its internal syntax and rules.

The SLH leads the corroborating probabilistic prediction hypothesis which Pearce states as:

while listening to new music, an enculturated listener applies models learned via the SLH to generate probabilistic predictions that enable them to organize and process their mental representations of the music and generate culturally appropriate responses. (Pearce, 2018). IDyOM works by providing the model with a musical corpus that it assumes is representative of a genre or musical style.

This musical corpus then serves as training data to approximate either a listener’s ground truth or musical style. After establishing this corpus, IDyOM then learns both long term and short term expectations of events using a variable-order Markov model in order to best optimize its predictive abilities in line with theoretical frameworks provided by @conklinMultipleViewpointSystems1995. The expectations that IDyOM calculates are based on a probability distribution of the proceeding events, which is then quantified in terms of information content [@shannonMathematicalTheoryCommunication1948]. As detailed in a summary review article on IDyOM by Pearce, IDyOM has been successful at predicting

Western listeners’ melodic pitch expectations in behavioral, physiological, and electroencephalography (EEG) studies using a range of experimental designs, including the probe-tone paradigm visually guided probe-tone paradigm a gambling paradigm, continuous expectedness ratings, and an implicit reaction-time task to judgments of timbral change. Peace notes some of IDyOM successes in modeling beyond expectation, including successes in modeling emotional experiences in music, recognition memory, perceptual similarity, phrase boundary perception and metrical inference.

Importantly in reviewing IDyOM’s capabilities regarding memory for musical pitches, Pearce also claims that

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

Peace notes four studies [@bartlettRecognitionTransposedMelodies1980; @cohenRecognitionTransposedTone1977; @cuddyMusicalPatternRecognition1981; @halpernAgingExperienceRecognition1995] that show that more complex melodies are more difficult to hold in memory. This theoretical assertion and select empirical findings have important ramifications for the aural skills classroom. In a dictation setting, melodies that are more expected should tax memory less, thus making them easier to remember and dictate. If I assume that more expected melodies are easier to remember, then it follows that the information content measures of expectedness can then be used as a stand in measure of melodic memory. This notion is not new to music psychology and was discussed by David Huron relating exposure to musical material as following similar laws to the the Hick-Hyman hypothesis [@hickRateGainInformation1952; @hymanStimulusInformationDeterminant1953], which Huron paraphrases as “processing of familiar stimuli is faster than processing of unfamiliar stimuli” [@huronSweetAnticipation2006, p. 63]. Now a decade later, this assertion can be further investigated using tools from computational musicology. Combining the Hick-Hyman hypothesis together with the above statistical learning hypothesis and probabilistic prediction hypothesis, I then put forward a new hypothesis: the frequency facilitation hypothesis.