LOOKING FOR TACTUS IN ALL THE WRONG PLACES: STATISTICAL INFERENCE OF METRIC ALIGNMENT IN RAP FLOW

Nathaniel Condit-Schultz
natcs@gatech.edu

ABSTRACT

Musical rhythm and meter are characterized by simple proportional relationships between event durations within pieces, making comparison of rhythms between *different* musical pieces a nebulous practice, especially at different tempos. Though the "main tempo," or tactus, of a piece serves as an important cognitive reference point, it is difficult to identify objectively. In this paper, I investigate how statistical regularities in rhythmic patterns can be used to determine how to compare pieces at different tempos, speculating that these regularities could relate to the perception of tactus. Using a Bayesian statistical approach, I model first-order (two-gram) rhythmic event transitions in a symbolic dataset of rap transcriptions (MCFlow), allowing the model to renotate the rhythmic values of each transcription as needed to optimize fit. The resulting model predicts makes "renotations" which match a priori predictions from the original dataset's transcriber. I then demonstrate that the model can be used to rhythmically align new data, giving an objective basis for rhythmic annotation decisions.

1. INTRODUCTION

Symbolic representations of music generally encode rhythm using integer-related *note-value* categories—whether expressed as durations or inter-onset-intervals (ioi). Absolute timing is encoded indirectly (if at all) as the *tempo* of a reference note-value, conventionally the quarter-note. The musical and psychological validity of this approach is well established, as the schematic syntax of musical rhythm is primarily determined by proportional relationships, not absolute (clock-time) durations [1]. [1] However, this approach also presents a problem: If only proportional relationships *within* a piece are rhythmically relevant, on what basis can relationships or comparisons be made across pieces? Can we be confident that a "quarter-note" in one piece is the same as a "quarter-note" in another? For example, consider three expert transcriptions of songs by Johnny Cash from the *RS200* dataset [3]: "Folsom Prison Blues" (1955) and "Ring Of Fire" (1963) are transcribed with quarter-notes at 110_{bpm} and 104_{bpm} respectively, while "I Walk The Line" (1956) is transcribed at 210_{bpm}. [2] These three songs share many idiomatic musical features, including *backbeat* strikes in between bass-notes at a 105–110_{bpm} pulse. Given these similarities, perhaps the quarter-notes in "I Walk the Line" ought to be compared to the eighth-notes in "Ring of Fire."

The quarter-note is more than a default reference unit for rhythmic encoding: It is also associated with the cognitive phenomenon of the "main beat" or *tactus*, and thus the "true" tempo of metric music [1, 4–6]. Other metric *levels* may be related to the tactus, both in notation and in human perception [1, 5, 6]. Thus, rhythmic comparison (in metric music) might be, essentially, a question of tactus comparison between two or more pieces. Which metric level in, for example, "Ring of Fire" or "I Walk the Line" is the tactus? This is essentially another perspective on the classic issue of "tempo octaves" in tempo-estimation research.

1.1 Background

Listeners must infer metric structure from music as they hear it [7], including the tactus level [5]. Though listeners' metric interpretations often agree [8], disagreement is also common, especially regarding tactus [4, 5, 8–11]. This suggests that tactus inference is constrained, but not determined, by features of music's objective organization. Which features constrain our perception of tactus? The obvious feature to consider is absolute (clock) time. Indeed, listeners tend to subdivide slower pulses or group faster pulses into beats in a preferred timing range, approximately corresponding to a tempo octave ($^2/_1$ ratio) of 160–80_{bpm} [5,10,12]. However, empirical measures of optimal tempo ranges have often covered larger ratios—from $^{2.25}/_1$ [12] to $^{2.5}/_1$ [13]—and a non-trivial number of observations spread across even more extreme tempos [4, 10]. Tempos from 200–60_{bpm} will feel somewhat familiar to most musicians, creating an "apparent contradiction between the narrow range of preferred tempi and the wide range of (absolute) tempi found in real music" [14]. These findings demonstrate that tactus perception is not determined by absolute timing in a trivial manner. Even if a strict tempo-octave were used for comparison, this still requires an arbitrary choice of the cutoff between tempo-octaves [10].

[1] In fact, human perception tends to normalize ioi ratios that are *not* simple ratios to the nearest simple-ratio category [2].

[2] "Ring of Fire" was transcribed by Temperley, the other two songs by de Clercq.

Attribution: F. Author, S. Author, and T. Author, "Looking for Tactus in All the Wrong Places: Statistical Inference of Metric Alignment in Rap Flow", in *Proc. of the 25th Int. Society for Music Information Retrieval Conf.,* San Francisco, United States, 2024.

Relative rhythmic features also contribute to tactus perception [4, 13–15]. In particular, the density and consistency of attacks at particular metric levels—what Martens [4] calls "pulse consistency"—serve as a important cue [6, 10]. Music theorists have also noted specific rhythmic patterns or aspects of the music's feel [3] that relate to tactus. A notable example is the backbeat pattern evident in the Johnny Cash examples above, which is often regarded as tactus defining [5, 17, 18]. However, De Clercq [17] has argued that absolute speed overwhelms the backbeat norm in many cases, and musical features must be balanced with absolute speed when inferring the tactus.

In traditions that rely on notated music, composers' explicit choice of note values and time signature might be regarded as the "correct" tactus; However, many scholars have noted that classical time signatures leave room for ambiguity regarding the true tactus [4,10,17,19,20]. Music from vernacular traditions pose an even more acute problem for research, as rhythm values must be chosen by a scribe [17]. If theory, convention, and intuition serve, we might hope that homogeneous collections of scores are coherently aligned. Unfortunately, representing metric alignment across pieces is not necessarily an important goal in traditions of music notation, and there are no clear standards for composers, transcribers, or arrangers to follow.

1.2 Hypothesis

If metric orientation is essential to the syntactic organization of music, then proper metric alignment of pieces is necessary to reveal structural similarities and generalize about rhythmic syntax in a body of music [17]. Conversely, any "misaligned" pieces—like "I Walk the Line," perhaps—add noise to empirical distributions and hinder musicological analysis. In this paper, I explore a novel statistical approach to aligning and comparing rhythmic patterns across pieces within stylistically homogeneous musical corpora. I hypothesize that regularities in proportionally-encoded rhythmic patterns can serve as consistent cues of metric alignment of pieces, independent of absolute speed. In other words, that specific rhythmic patterns or features (notably, pulse saliency) will be statistically associated with particular metric levels, and that these patterns can then be used as the basis to align and compare pieces. To achieve this, we can systematically rescale note values of transcriptions—either in *augmentation* (longer values) or *diminution* (shorter values)—so as to optimize the fit of statistics related to syntactic rhythm relationships. For example, we could renotate "I Walk the Line" in diminution, and then confirm if the resulting tabulation of the overall RS200 collection is less noisy, "expos[ing] connections that would be otherwise hidden or obscured" [17]. My argument is that these connections, should they be revealed, may relate to listeners' perceptual experience of the tactus, though I cannot directly demonstrate that here.

A central premise of my hypothesis, is that metric alignment can be done based on proportional rhythmic data, without absolute timing information. This does not preclude that absolute timing plays an important role in musical alignment, but if the hypothesis is supported, it would demonstrate that rhythmic syntax is at least partly independent of tempo, and help explain why tempos are used outside a preferred tempo octave.

2. METHODOLOGY

With no ground truth available, I can only attempt to optimize fit to my data in an unsupervised way. My approach is to characterize empirical probability distributions of rhythmic data conditioned on different interpretations of the metric alignment of pieces.

2.1 Data

For this project, I use my own Musical Corpus of Flow (MCFlow) [18], in which I transcribed the rapped part of 124 popular hip-hop songs, all in $\frac{4}{4}$ time. Rap flow is suitable for this task for several reasons: Rap flow is saturated with the rhythmic features of American popular music more broadly, with lots of rhythmic variation within songs. Rap also tends to exhibit a rhythmically dense, fast pace, with few long iois, which makes relatively simple ngram-like analyses (described below) more plausible. I parsed the MCFlow dataset using humdrumR [21]—a R package for analyzing data encoded in the humdrum syntax (as MCFlow is). I restrict my analysis to inter-stress-intervals because rap scholars agree that most useful rhythmic information is in the stressed syllables of rap [18, 22].

MCFlow divides each of its 124 songs into verses. In some cases, different artists perform different verses, occasionally even at different tempos. I thus regard each verse as a separate rhythmic passage to analyze. To isolate only "pure" duple rhythmic data, I remove 155 measures of music, in 44 unique verses, which contain at least one triplet. I then removed 16 verses with fewer than eight measures remaining. This leaves a total of 376 verses, containing 36,553 stressed syllables; the shortest remaining verse has only 21 stressed syllables, with the longest containing 314 and a median length of 98.

In my MCFlow transcriptions, I used the backbeat in the rap's accompaniment to determine the quarter-note value [18]. However, one of the most important reasons I use MCFlow is because I [23] originally noted thirty-five verses (in eleven [4] songs) which are clear outliers in tempo annotation (Table 1), and speculated that they might be better notated at a different tempo. This gives us a set of *a priori* predictions about metric alignment in the data. Figure 1 illustrates the distribution of quarter- and eighth-note iois in MCFlow, as notated (above) and incorporating my speculated renotations (below). Interestingly, the

[3] Another plausible area where musical organization might influence metric alignment is sub-syntactic *micro-timing*: small discrepancies between actual rhythmic timing and their perceived rational categories. Micro-timing is often related to the "feel" of music, and can be used to emphasize particular beat levels [16].

[4] I also identified two other outlier verses which I exclude because they contained fewer than eight measures of non-triplet bars.

Song	Verse(s)	BPM
Dead and Gone (T.I., 2009)	1–2	68
Niggas in Paris (Jay Z and Kanye West, 2011)	1–3	70
Mercy (Kanye West, at al., 2012)	1,2,4	70
What's Your Fantasy (Ludacris, 2000)	1–3	70
Holy Grail (Jay Z, 2013)	1–2	72
How Low (Ludacris, 2009)	1–2	72
Woof (Snoop Dogg, 1998)	1–3	83
Pray (M.C. Hammer, 1990)	1–5	122
It's Tricky (Run-D.M.C., 1987)	1–4	128
You Be Illin' (Run-D.M.C., 1986)	1–4	128
Fight for Your Right (the Beastie Boys, 1987)	1–3	134
Mercy (Kanye West, at al., 2012)	3	140

Table 1. List of verses in MCFlow which Condit-Schultz [18] identified as tempo outliers. BPM = quarter-notes per minute.

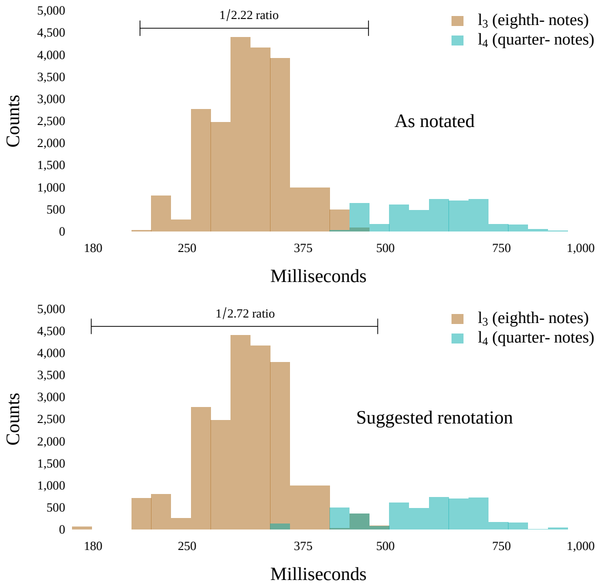

Figure 1. Distribution of notated quarter- and eighth-note inter-stress-intervals in MCFlow, by absolute duration.

raw, backbeat-based notation covers an tempo range only slightly greater than one tempo octave, similar to observed listener preferences [12, 13]. In contrast, following my speculated renotations results in a few verses being moved into more extreme absolute tempos.

2.2 Meter

Meter is an organizational structure in music, wherein multiple phase-aligned beats with integer-related periods form a nested hierarchical pattern [5]. These beats can be sorted from fastest ("lowest") to slowest ("highest"), each considered one *metric level*, notated here as $[l_1, \ldots, l_k]$. The highest metric level (l_k) defines the overall period of the meter, a *measure*; the lowest metric level (l_1) is known as the *tatum*. In a musical passage, each note onset is associated with a tatum pulse [24], and thus a unique *metric position* within each measure—e.g., "beat 4." Metric positions may also coincide with one or more higher-metric levels, with the highest level defining the "level" of that position. For example, the downbeat of each measure is the unique position at level l_k.

In this paper, I consider only simple duple meter, with each metric level having twice the period of the level below it [1]: essentially a $\frac{4}{4}$ meter with strictly no triplets. The standard $\frac{4}{4}$ generally presumes at least three central levels [5, 20]. However, music often evinces hyper-metric pulses above the measure level [20] and, conversely, faster levels well below the ostensbile tactus (e.g., 16th- and 32nd-notes). Thus, I proceed with a slightly expansive $k = 6$: six metric levels with 32 metric positions. This could be interpreted as one measure of 32nd-notes, two measures of sixteenth-notes, or four measures of eighth-notes. Throughout this paper, I will take l_1 as 32nd-notes, putting quarter-notes in l_4.

Regardless of notation, the fastest metric level (tatum) can always be identified in any transcription. However, some musical passages may have *implicit* subdivisions, that would be felt by a listener, but are never articulated in the music. Thus, the *true* tatum l_1 may be different than the observed tatum $l_{\bar{1}}$. For any given musical transcription, we can postulate one or more implicit subdivisions, effectively "shifting" the observed metric positions up one level—equivalent to renotating the music in augmentation.

2.2.1 *Modeling Meter*

To characterize the rhythms of music in metric terms, I use a first-order (two-gram) model, considering the joint probability of the metric positions of sequential pairs (antecedent-consequent) of note events. Given 32 positions, a full transition matrix would require 1,024 parameters, many of which would be close to zero or simply redundant, as rhythmic patterns in different parts of the measure can be closely related. To work with less sparse and more interpretable parameters, I explored ways of reducing the full 32x32 parameters space to a smaller number of parameters while maintaining predictive power. My final approach is to bin each antecedent note according to its metric level l_k and each consequent into one of nine categories defined relative to the antecedent position. My nine *metric consequent types*, illustrated in Figure 2, are able to differentiate between shorter and longer iois, weak-to-strong versus strong-to-weak beat transitions, and different sorts of syncopations. These forty-seven parameters are represented in a vector $\mathbf{p}$, with components $p_{l.m}$ corresponding to the probability of each metric transition (Table 2). In the raw MCFlow data, the full 32X32 metric transition matrix has a joint entropy of 6.19 bits (10 being the maximum theoretical value). My antecedent-consequent parameterization, with only 47 parameters, achieves a cross-entropy with the same data of 7.44 bits, gaining only 1.25 bits [5] by removing 977 parameters. Figure 3 illustrates the distribution of my antecedent-consequent parameters, using the raw MCFlow note-values.

2.3 Statistical Model

The statistical model I employ mirrors the thought process we explored with Johnny Cash songs above, "renotating"

[5] This difference in bits is equivalent to the Kullback–Leibler divergence.

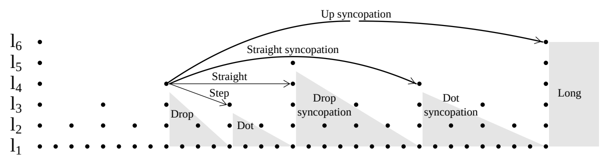

Figure 2. Illustration of nine metric consequent types at the l_4 level. Arrows point to exact points; shaded areas indicate points binned together.

	Drop	**Step**	**Dot**	**Straight**	**Drop sync.**	**Straight sync.**	**Dot sync.**	**Up sync.**	**Long**
l_6	$p_{6.1}$	$p_{6.2}$	$p_{6.3}$	$p_{6.4}$	$p_{6.5}$		$p_{6.7}$		$p_{6.9}$
l_5	$p_{5.1}$	$p_{5.2}$	$p_{5.3}$	$p_{5.4}$	$p_{5.5}$	$p_{5.6}$	$p_{5.7}$	$p_{5.8}$	$p_{5.9}$
l_4	$p_{4.1}$	$p_{4.2}$	$p_{4.3}$	$p_{4.4}$	$p_{4.5}$	$p_{4.6}$	$p_{4.7}$	$p_{4.8}$	$p_{4.9}$
l_3	$p_{3.1}$	$p_{3.2}$	$p_{3.3}$	$p_{3.4}$	$p_{3.5}$	$p_{3.6}$	$p_{3.7}$	$p_{3.8}$	$p_{3.9}$
l_2	$p_{2.1}$	$p_{2.2}$	$p_{2.3}$	$p_{2.4}$	$p_{2.5}$	$p_{2.6}$	$p_{2.7}$	$p_{2.8}$	$p_{2.9}$
l_1				$p_{1.4}$		$p_{1.6}$		$p_{1.8}$	$p_{1.9}$

Table 2. Forty-Seven metric coefficients ($\mathbf{p}$). Empty slots are logically impossible given the definitions. sync. = syncopation.

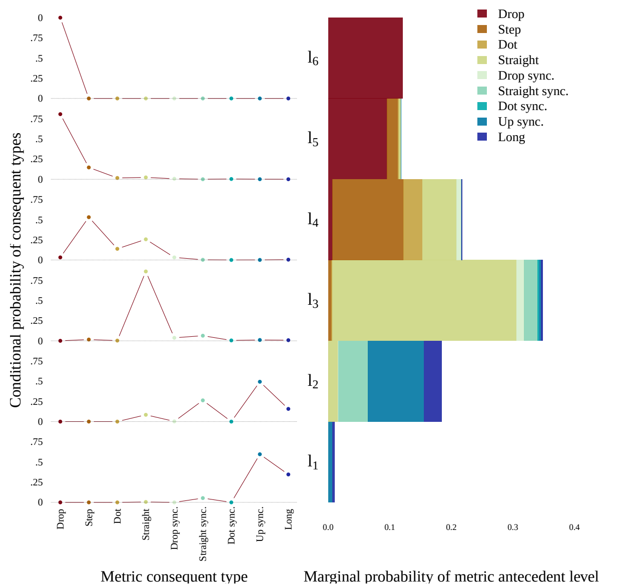

Figure 3. Raw empirical estimates for $p_{l.m}$. Both sides of the figure show the same information in two different formats: the left side shows the conditional probability of each metric consequent, given the metric level of an antecedent syllable; the right side shows the joint probability of the same antecedent-consequent pairings.

verses in the MCFlow corpus to find a good fit. With each iteration of the model's Monte Carlo algorithm, the model finds estimates of the metric coefficients (explained below) using the dataset as currently encoded. The model then estimates parameters which represent the "scaling" of *each* individual verse, by retabulating the music assuming one or two unobserved sub-divisions—equivalent to renotating the music in augmentation. The process repeats, reestimating the meter parameters using the new scaling parameters, etc., until a complete picture of the posterior distribution emerges, as guaranteed by the Metropolis-Hasting algorithm [25]. Ultimately, I find the scalings of each verse that result in the best fit to the overall metric distribution.

In each verse, I count instances of 47 metric transition bins, indexed $l.m$ as defined above for $\mathbf{p}$. Let the counts in the nth verse be labeled $\mathbf{C}^n = [c^n_{l.m}, \ldots]$. I then model each set of counts as an independent draw from a multinomial distribution $C^n \sim \mathcal{M}(\sum C^n, \mathbf{p})$. The core purpose of the project, however, is to estimate a set of indicator, "shift," parameters, one for each verse: $\mathbf{s} = [s^1, \ldots, s^n]$, where $s^n \in \{0, 1, 2\}$. There are thus actually three different counts ($C^{n(s \in \{0,1,2\})}$) for each verse, one for each possible shift parameter: When $C^{n(s=0)}$, the metric parameters are counted assuming the observed tatum is the true tatum $l_{\bar{1}} = l_1$. When $C^{n(s=1)}$, count assuming that there is one implicit level of duple subdivision in the meter, $l_{\bar{1}} = l_2$, "shifting" the metric parameters up one level. When $C^{n(s=2)}$, counted assuming two subdivisions, $l_{bar1} = l_3$. I assume also that the values of $\mathbf{s} \sim \mathcal{M}(n, \mathbf{S})$, where $\mathbf{S} = [S_0, S_1, S_2]$ is another discrete probability distribution (though this ultimately had little impact on my results).

2.3.1 *Model Estimation*

Given the assumed distributions above, I use a Bayesian Markov Chain Monte Carlo (MCMC) algorithm to calculate posterior distributions for $\mathbf{p}$, $\mathbf{s}$, and $\mathbf{S}$. Since objective estimates for the $\mathbf{s}$ shifting parameters are my main goal, I specify no prior distribution for $\mathbf{s}$, letting the model believe (initially) that all values of $\mathbf{s}$ are equally probable. For $\mathbf{p}$ and $\mathbf{S}$, I specify minimally informed Dirichlet prior distributions: $prior(\mathbf{p}) \sim \mathcal{Dir}(\alpha_{l.m} = 5)$ and $prior(\mathbf{S}) \sim \mathcal{Dir}(\alpha_S = 5)$. These minimal priors—equivalent to observing 235 prior note transitions and 15 prior verse shifts respectively—mainly serve to (weakly) discourage the model from assigning values close to zero. Note that the Bayesian approach here does not only find the optimal point-estimate for each parameter, but a complete prior distribution of belief regarding each parameter. This will allow the model to express degrees of certainty about each s^n, rather than finding only one optimal choice.

I estimate the posterior distribution using a custom MCMC implementation in R, with three Gibbs-sampler steps (for $\mathbf{s}$, $\mathbf{S}$, and $\mathbf{p}$) in each iteration, i. In each Gibbs step, I sample new parameter estimates for one parameter from the conditional distribution of that parameter given the current values of the other parameters. The result is a sequence of estimates for each parameter, forming a

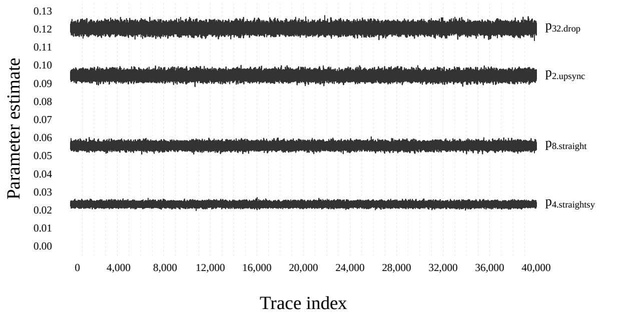

Figure 4. MCMC trace for four selected parameters. Dashed lines indicate boundaries between independent chains.

Markov chain which converges on the true posterior distribution.

For the s scaling parameters, new estimates for all $[s^{1\dots n}]$ parameters are sampled in a single step. For each verse, the probability of observing $C^{n(s)}$ for all three values of s, conditioned on $\mathbf{p}_i$ and $\mathbf{S}_i$, is computed.

$$s_{i+1}^n \sim P(C^{n(s)}|\mathcal{M}(\sum_1^n C_i^{n(s)}, \mathbf{p_i})) * \mathbf{S}_i$$

For the $\mathbf{p}$ metric coefficient parameters, new estimates for all parameters were sampled in a single step, conditional only on $\mathbf{s_i}$. Taking advantage of the conjugate relationship between the multinomial and Dirichlet distributions, I can sample from the conditional distribution of $\mathbf{p}$ directly using the Dirichlet distribution:

$$\mathbf{p}_{i+1} \sim \mathcal{D}ir(\alpha = \sum_{n=1}^{N} C^{n(s_i)}{}_i + prior_\alpha(\mathbf{p}))$$

Updates for $\mathbf{S}$ are similar but even simpler, using only the current (estimated) counts of $\mathbf{s}$: $\mathbf{S}_{i+1} \sim \mathcal{D}ir(\alpha = counts(\mathbf{s}_i) + prior_\alpha(\mathbf{S}))$.

To minimize the effect of initial values, I initialized forty independent markov chains on different random draws from the prior distributions of $\mathbf{p}$ and $\mathbf{S}$, and a uniform random sample of $\mathbf{s}$ parameters. Each chain ran for 11,000 samples, with an initial "burn in" of 1,000 iterations removed from each chain, though each chain appeared to reach its stationary distribution well before the 1,000th iteration. All forty chains converged on the same final distributions for all parameters (see Figure 4 for a few examples). As is usually the case with MCMC models, several parameter chains evinced moderate autocorrelation values (the largest being 0.287), so I thinned the chain by taking every tenth sample, cutting the absolute autocorrelation values down to $r \leq 0.115$. The result is a chain of 40,000 samples for each parameter. Figure 4 shows the MCMC traces for four of the $\mathbf{p}$ parameters; the other parameter traces look essentially identical.

3. RESULTS

The main parameters I am interested in are the estimates of $\mathbf{s}$, the "shift" parameters for each verse. Despite the fairly long MCMC trace (40,000 samples), in 372 of 376 verses the model selected the same shift parameter in *every* sample; evidently, most verses fit in one, and only one, interpretation. In only three verses—none of which were *a priori* tempo outliers—did the model find significant uncertainty, with the non-modal choice sampled between 12.6% and 43.9% of the time (these appear midway between shift levels in Figure 5). The important question is whether these highly confident shift parameters match my *a priori* expectations. If we take the posterior modal value for each s^n, we observe 37 shifts of 0, 318 of 1, and 21 of 2. Figure 5 shows these average posterior $\mathbf{s}$ values normalized relative to the original empirical $l_{\bar{1}}$ of each verse, such that 0 indicates the original notated quarter-note. The model correctly identifies the predicted renotation for 27 of the 35 *a priori* outliers. The model also identifies four unanticipated verses that need shifting, and fails to shift eight verses—if we view this as a binary classification task, the model achieves an F-score of .818. Note that the model was not provided any information about absolute timing, so this accuracy is achieved purely by looking at metric transitions. Close investigation of the false positives reveals that, though I didn't originally identify these verses as outliers [18], each features flow that could make sense renotated. The false negatives are not as easy to interpret; However, in no case did the model falsely reject all outlier verses in a song: for example, the model correctly shifts four of the five verses in MC Hammer's "Pray," but fails to shift the fourth verse (for no obvious reason).

To visualize the posterior distribution of $\mathbf{p}$, the metric coefficients, I show the average posterior value for each $p_{l.m}$ parameter in Figure 6. If we compare this to Figure 3, there are no dramatic differences, except at l_1, where the model places considerably less probability mass. The average entropy of the $\mathbf{p}$ posterior is 3.339 bits, slightly lower than the joint distribution of the raw-notation counts at 3.409 bits, demonstrating that model has improved the overall fit of the data. Finally, I can also use the posterior parameters to evaluate unseen data. For example, if I apply the posterior $\mathbf{p}$ to our three Johnny Cash songs, we find that the model shows (with total confidence) that the three songs *should* be aligned at the same backbeat level, as I speculated at the outset.

4. DISCUSSION AND CONCLUSION

Though it appears that my statistical approach both improves fit and matches (my) expert judgments [18], this initial foray is not a decisive demonstration that this approach can help us generalize about metric syntax. It appears that my model is learning, or at least observing, *something* about the organization of metric syntax across metric levels, but future work is needed to elucidate what is going on, and determine how robust this methodology can be.

It may seem that my full Bayesian treatment is overkill:

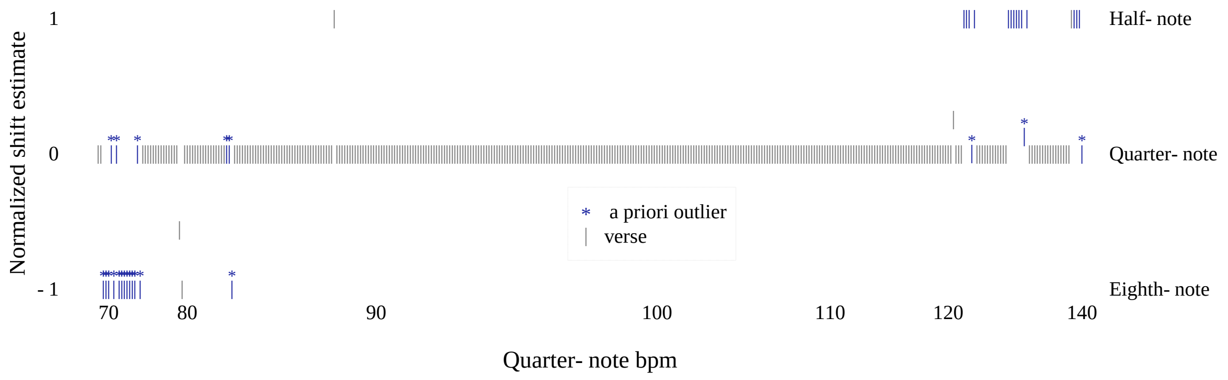

Figure 5. Average posterior **s** estimates for each verse, sorted by raw quarter-note bpm. Blue marks indicate the predicted tempo outliers. The shifts have been "normalized" such that 0 indicates the original notated quarter-note, rather than the fastest metric level.

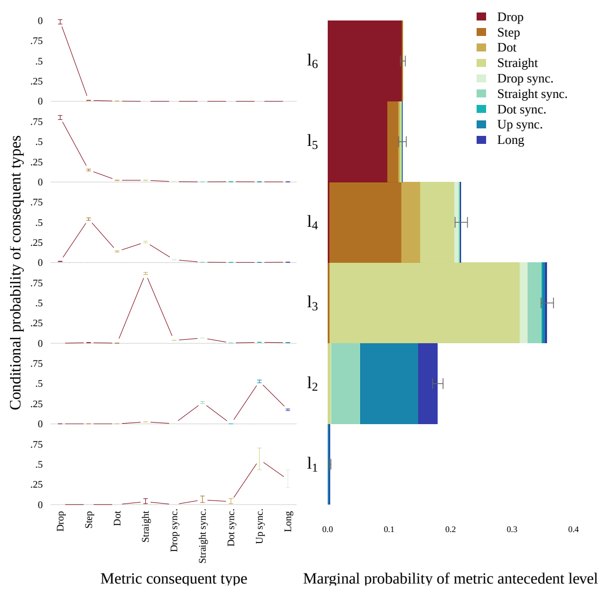

Figure 6. Mean posterior parameter estimates for $p_{l.m}$. Both sides of the figure show the same information in two different formats: the left side shows the conditional probability of each metric consequent, given the metric level of an antecedent syllable; the right side shows the joint probability of the same antecedent-consequent pairings. Bars indicate the Bayesian 95% credible interval for each parameter.

My parameter estimates are tightly packed around their mean and not dramatically different than the simple counts derived from the raw data; my posterior estimates of **s** also show little variability. This suggests that a simpler approach could probably achieve similar results on this dataset. My results are also strongly fitted to this particular dataset—for this initial attempt I specified uninformative priors on all parameters, allowing the model to fit the data at hand very closely. However, I believe this full Bayesian approach will prove robust if extended to other datasets which might be less rhythmically uniform than MCFlow, and the results here could be used as the basis for more informative priors for future work.

Finally, though I have argued that this task is theoretically connected to perceptual and musicological ideas of tactus and tempo, future work with human participants will be necessary to establish direct connections between my findings and human perception. For example, my **p** estimates could be used to generate rhythmic stimuli with different (predicted) tactus interpretations. For course, as discussed above, there is plenty of evidence that tactus is never fully determined by musical features [4, 5, 8–11]. Listeners' perception might be shaped previous context, personal experience, their own personal state, or by conscious effort. My analysis of "syntactic regularities," even if valid, isn't necessarily connected to tactus at all: indeed, at least one prominent psychomusicological theory of rhythm, London's [20] (p. 95) *tempo-metrical types*, "is [explicitly not] defined in terms of the level heard as the tactus." It's possible that the statistical regularities found by my model represent tempo-metrical types, or other rhythmic structural principles, but *not* tactus.

Basing psychological conclusions on statistical evidence requires a match between the musical corpora and the listening experience of people. Different musical exposure (and thus statistical experience) might explain disagreements about tactus. Stepping back further, it is possible that syntactic relationships in music involve relationships between various rhythms and beats *without* assuming any privileged reference level at all. My results here make this final possibility appear unlikely, but much work remains to be done.

5. REFERENCES

[1] M. R. Jones and M. G. Boltz, "Dynamic attending and responses to time," *Psychological Review*, vol. 96, no. 3, p. 459–491, 1989.

[2] N. Jacoby and J. H. McDermott, "Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction," *Current Biology*, vol. 27, no. 3, p. 359–370, 2017.

[3] D. Temperley and D. Tan, "Emotional connotations of diatonic modes," *Music Perception*, vol. 30, no. 3, p. 237–257, February 2013.

[4] P. A. Martens, "The ambiguous tactus: Tempo, subdivision benefit, and three listener strategies," *Music Perception*, vol. 28, no. 5, p. 433–448, June 2011.

[5] P. A. Martens and F. Benadon, "Musical structure: Time and rhythm," in *The Routledge Companion to Music Cognition*, R. Ashley and R. Timmers, Eds. Taylor & Francis, 2017, p. 115–127. [Online]. Available: https://doi.org/10.4324/9781315194738

[6] R. Parncutt, "A perceptual model of pulse salience and metrical accent in musical rhythms," *Music Perception*, vol. 11, no. 4, p. 409–464, July 1994.

[7] H. C. Longuet-Higgins and C. S. Lee, "The rhythmic interpretation of monophonic music," *Music Perception*, vol. 1, no. 4, p. 424–441, July 1984.

[8] C. K. Madsen, R. A. Duke, and J. M. Geringer, "The effect of speed alterations on tempo note selection," *Journal of Research in Music Education*, vol. 34, no. 2, p. 101–110, 1986. [Online]. Available: http://www.jstor.org/stable/3344738

[9] D. Hammerschmidt and C. Wöllner, "Sensorimotor synchronization with higher metrical levels in music shortens perceived time," *Music Perception*, vol. 37, no. 4, p. 263–277, March 2020.

[10] M. F. McKinney and D. Moelants, "Ambiguity in tempo perception: What draws listeners to different metrical levels?" *Music Perception*, vol. 24, no. 2, p. 155–166, December 2006.

[11] P. Toiviainen and J. S. Snyder, "Tapping to bach: Resonance-based modeling of pulse," *Music Perception*, vol. 21, no. 1, p. 43–80, September 2003.

[12] L. van Noorden and D. Moelants, "Resonance in the perception of musical pulse," *the Journal of New Music Research*, vol. 28, no. 1, p. 43–66, 1999.

[13] S. Quinn and R. Watt, "The perception of tempo in music," *Perception*, vol. 35, no. 2, p. 267–280, 2006, pMID: 16583770. [Online]. Available: https://doi.org/10.1068/p5353

[14] G. Madison and J. Paulin, "Ratings of speed in real music as a function of both original and manipulated beat tempo." *the Journal of the Acoustical Society of America*, vol. 128, no. 5, p. 3032–3040, 2010. [Online]. Available: http://search.proquest.com/docview/815547411/

[15] J. London, "Tactus $\neq$ tempo: Some dissociations between attentional focus, motor behavior, and tempo judgment," *Empirical Musicology Review*, vol. 6, no. 1, p. 43–55, 2011.

[16] C. Drake, A. Penel, and E. Bigand, "Tapping in time with mechanically and expressively performed music," *Music Perception*, vol. 18, no. 1, p. 1–23, October 2000.

[17] T. de Clercq, "Measuring a measure: Absolute time as a factor for determining bar lengths and meter in pop/rock music," *Music Theory Online*, vol. 22, no. 3, 2016. [Online]. Available: https://doi.org/10.30535/MTO.22.3.3

[18] N. Condit-Schultz, "MCFlow: A Digital Corpus of Rap Transcriptions," *Empirical Musicology Review*, vol. 11, no. 2, p. 124–147, 2016.

[19] M. R. Jones, R. R. Fay, and A. N. Popper, Eds., *Music Perception*, ser. Springer Handbook of Auditory Research. Springer, 2010, vol. 26.

[20] J. London, *Hearing in Time: Psychological aspects of musical meter*, 2nd ed. Oxford: Oxford University Press, 2012.

[21] N. Condit-Schultz and C. Arthur, "humdrumr: a new take on an old approach to computational musicology," in *Proceedings of the International Society for Music Information Retrieval*, November 2019, p. 715–722.

[22] M. Ohriner, *Flow: The rhythmic voice in rap music*. Oxford Studies in Music Theory, 2019.

[23] N. Condit-Schultz, "MCFlow: A Digital Corpus of Rap Flow," Ph.D. dissertation, Ohio State University, 2016.

[24] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*. Cambridge, Massachusetts: The MIT Press, 1983.

[25] S. Chib and E. Greenberg, "Understanding the metropolis-hastings algorithm," *The American Statistician*, vol. 49, no. 4, pp. 327–335, 1995. [Online]. Available: http://www.jstor.org/stable/2684568