Evaluating and Extending
Computational Models of Rhythmic Syncopation in Music[*]

Leigh M. Smith and Henkjan Honing
Music Cognition Group
Institute for Logic, Language and Computation, Universiteit van Amsterdam
lsmith@science.uva.nl, http://www.hum.uva.nl/mmm

Abstract

What makes a rhythm interesting, or even exciting to listeners? While in the literature a wide range of definitions of syncopation exists, few allow for a precise formalization. An exception is Longuet-Higgins and Lee (1984), that proposes a formal definition of syncopation. Interestingly, this model has never been challenged or empirically validated. In this paper the predictions made by this model, along with alternative definitions of metric salience, are compared to existing empirical data consisting of listener ratings on rhythmic complexity. While correlated, noticable outliers suggest processes in addition to syncopation contribute to listeners judgements of complexity.

1 Musical Surprise

The experience of surprise in listening to music can be considered as a necessary, although not sufficient, inverse indicator of the listeners expectation. In addition to confounding expectations from pitch and harmony, surprise arises temporally, contributed notably by metrical salience, the presence and role of an anacrusis, expressive timing (for example *tempo rubato* and asynchrony), and syncopation. The musical concept of syncopation is well known to most performers of Western music, regularly used by composers and commonly taught in music education.

Recently, there has been significant research in musical expectancy, using formal models (see Smith, 1999, for a review). As distinguished by Bharucha (1993, pp. 498), musical expectation can be *schematic* — abstract knowledge derived over time and many examples — or *veridical*, derived from the implications of the particular musical events attended to during a performance. The skillful interplay between expected and surprising events is critical to the generation of musical expression (Meyer, 1956). Robust and explicit management of expectation and of surprise has a computational value also, to enable artificial performers or tools to fail as gracefully as humans do, both when accompanying or interpreting human performers and in the feedback loop evaluating these systems performance of some hardware or software instrument.

The approach of connectionist recurrent networks (Bharucha and Todd, 1991) and hidden Markov models (HMM) is to encode the weight of tendency to transition from one note to the next given previous history. These approaches learn expected transitions by training with musical examples. A key issue with these artificial systems is their degree of explanatory function in cognitive and musicological terms, otherwise the explanation can only be by examining and interpreting the weights located in the HMM matrix or neural net connection strengths. This casts the notion of surprise only in terms of transitions of notes. Such systems do not seem to adequately account for schematic expectancy: How does a learning algorithm resolve a veridically surprising yet schematically expected event? Will training these systems on sufficiently large sets of examples properly create schematic learning, or lead to over-fitting to training sets? We therefore first investigate formal rule-based, symbolic models.

2 Syncopation

There are few examples of formal representation of syncopation. An exception is Longuet-Higgins and Lee (1984) (LH&L), that assumes the listener will attempt to interpret a rhythm according to a given meter so as to minimize syncopations. Syncopation is consequently defined by them as a beat stronger than the previous sounded note falling on a rest or tied note (see Figure 1). A syncopation occurs if and only if a (sounded) note outlasts the highest-level *metrical unit* it
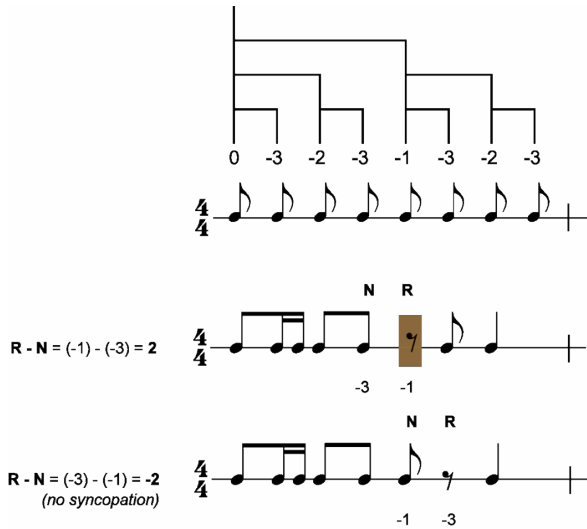
Figure 1: LH&L syncopation measure calculated from a theoretical metric salience tree (top diagram). Syncopation is calculated as the difference between the initiating metrical unit of the rest (R) and it's preceding sounding note (N).

initiates. Metrical units are defined in a hierarchy identical to Lerdahl and Jackendoff's (1983) metrical hierarchy, although inverse in polarity. Meters in the LH&L model are specified in terms of a vector of rhythmic subdivisions M, for example $\frac{6}{8}$ is represented by (2 3), subdividing the measure into two parts, and then each of those into three.

Palmer and Krumhansl (1990) argue that pre-established mental frameworks ("schemas") for meter are used during listening. These schemas enable robust interpretation despite sometimes contradicting, ambiguous or absent objective cues. They hypothesize that "Strong statistical regularities for meter may allow emphasis of other musical dimensions such as pitch or intensity by reducing the attention listeners must allocate to meter and increasing listeners anticipations for certain event locations" (Palmer and Krumhansl, 1990, pp. 733).

In this study they tested the types of mental structures for meter ("perceptual hierarchies") evoked from simple event sequences. Listeners rated the relative acceptability of audible events at different locations in the metrical grid (Palmer and Krumhansl, 1990, Experiment 2). They found a significant difference in performance between musicians and non-musicians, arguing that musicians hold more resilient representations of meter, which favours hierarchical subdivision of the measure, than the non-musicians. These two ratings can be interpreted as an alternative to the LH&L metric saliences ($S_{LH\&L}$), referred here as $S_{P\&K-M}$ and $S_{P\&K-NM}$, and derived in Section 3.

3   Method

To aid comparison against multiple examples and models, the LH&L measure of syncopation was implemented in Common Lisp. This allows for the substitution of alternative measures of metrical salience. To substitute the P&K metric salience measures for the LH&L hierarchies required the following additions to the model:

1. The appropriate metrical hierarchy vector H must be selected based on the lookup of a dictionary containing perceptual hierarchies keyed by a "canonical" meter (fully specified to a maximum number of 12 or 16 time subdivisions) and subject expertise (musician vs. non-musician). Incompletely specified meters are first compared to the minimum prefix of the canonical versions. The matched canonical version then is used to index the 12 or 16 element vector H.

2. The P&K perceptual hierarchies must be scaled down to match the LH&L metrical hierarchy polarity and magnitude:

$$S = \left[\frac{H - \min(H)}{\max(H) - \min(H)} - 1 \right] * d \qquad (1)$$

where S is the metrical salience vector produced ($S_{P\&K-M}$ or $S_{P\&K-NM}$), H is the perceptual hierarchy vector, d is the depth of the metrical hierarchy, that is, the number of subdivisions of the meter, being the length of the meter divisor list M.

3. Select every nth item of the retrieved and scaled perceptual hierarchy to match the meter specification, where

$$n = \frac{len(H)}{\prod M}. \qquad (2)$$

For example M = (2 2 2) matches against the canonical meter of (2 2 2 2). This selects the $\frac{4}{4}$ meter perceptual hierarchy H. Meter (2 2 2) selects 8 of the originally recorded hierarchy of 16 ratings, every second rating.

The model is applied by

$$N_R = syncopate(S, M, R) \qquad (3)$$

where S is the metrical salience, either determined by LH&L's original metrical salience algorithm, or Equation 1 for P&K's model, M is the meter divisor list, R is the test rhythm. The Longuet-Higgins and Lee $syncopate()$ function produces a vector of syncopations $n_x \in N_R$, where-ever in the rhythm R a syncopation occurs. In order to compute a single syncopation measure for the entire rhythm, the syncopation measures were simply summed:

$$\sigma_R = \sum_{x=0}^{len(N_R)} n_x. \qquad (4)$$

This models the assumption that the entire experience of syncopation is created by it's contributing occurrences within the rhythm.

3.1 Evaluation of Metric Salience Measures

Three alternative measures of metric salience were tested. The first $S_{LH\&L}$, being the original metric salience, and two alternatives described from P&K, $S_{P\&K-M}$ and $S_{P\&K-NM}$. The predictions made by these three versions of the model were tested against a set of 35 rhythms that were judged for complexity on a scale of 1 to 5 (least to most complex, Shmulevich and Povel, 2000). Since all of the rated rhythms were 16 elements in length, the $\frac{4}{4}$ and $\frac{2}{4}$ perceptual hierarchies of Palmer and Krumhansl were tested. In order to compare these results, the output from LH&L and P&K models were linearly scaled to match the listener ratings of complexity.

4 Results

The comparison of the models to Shmulevich and Povel's listener ratings are shown in Figure 3 and Figure 4 and the models are compared in Figure 2. The P&K $\frac{4}{4}$ demonstrated better correlation with the listener ratings, and the stimulus from which that metric salience was derived better matched the inter-onset intervals of the listener ratings than $\frac{2}{4}$, so only the $\frac{4}{4}$ hierarchy is plotted in the figures.

The predictions made by the LH&L and P&K musician models positively correlate to the Shmulevich and Povel listener ratings ($r = 0.75$ and $r = 0.73$ respectively, both $p < 0.001$ (Press et al., 2002, pp. 641)).

The root mean square error (RMSE) goodness-of-fit between each prediction and the listener ratings were also similar (LH&L 0.75, P&K musician 0.76, lower values indicates greater goodness-of-fit). However, even if a high goodness-of-fit on one model had been achieved using RMSE, or percentage of variance accounted for (PVAF) measures, these alone are insufficient to select a model, since a good fit may actually indicate over-fitting. Such measures are unable to distinguish between variations in the data caused by noise, and those that the model is designed to capture (Honing, 2006).

Correlation of the P&K non-musician model to the listener ratings was less than the P&K musician model ($r = 0.64$, $p < 0.001$). This difference between correlation measures of the P&K non-musician model and the LH&L model could not be determined to be statistically significant, even
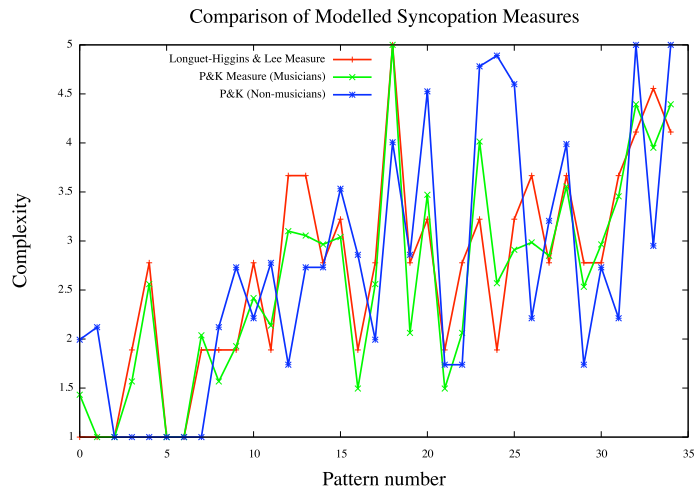

Figure 2: Comparison of the predictions of the models based on the metric saliences $S_{LH\&L}$, $S_{P\&K-M}$ and $S_{P\&K-NM}$. The performance of models are compared using Shmulevich and Povel's 35 example rhythms. The predictions of the models have been scaled to match the listener ratings on a scale of 1–5 (least to most complex).

assuming binormality between the sets (Press et al., 2002, pp. 642).

The tests highlight two categories of rhythms of interest, those that agree between model predictions, but differ between models and observed data, and those that differed between model predictions. Examples of both are rhythms 5^1 and 6^2 which were rated by listeners as marginally complex (2.08, 2.12 respectively), yet in the classical LH&L model, and both musician and non-musician P&K models, there is no syncopation anywhere in the measure. Rhythm 18,[3] rated moderately complex by listeners (2.84) produced the maximum syncopation measures for LH&L and P&K musician, and a substantial measure for the P&K non-musician. This is possibly from listeners interpreting the rhythm as a compound rather than syncopated hierarchical meter. Conversely, rhythm 24^4 produced low but differing syncopation predictions for LH&L and P&K musician models, and a high syncopation measure for P&K non-musician model.

Such variation indicates how critical the metrical saliences are to the performance of the models. Therefore using different saliences at specific metrical positions should produce a better estimate of perceived complexity. However, using $S_{P\&K-M}$ and $S_{P\&K-NM}$ did not produce better results, the performance using $S_{P\&K-M}$ was actually slightly less than

---

[1]Rhythm from Shmulevich and Povel (2000): | . | | | . | | | . . | | . . . |

[2] | | | . | . | | | . . | | . . . |

[3] | | | | . | . | . . | | | . . . |

[4] | | | . . | | . | | | . | . . . |

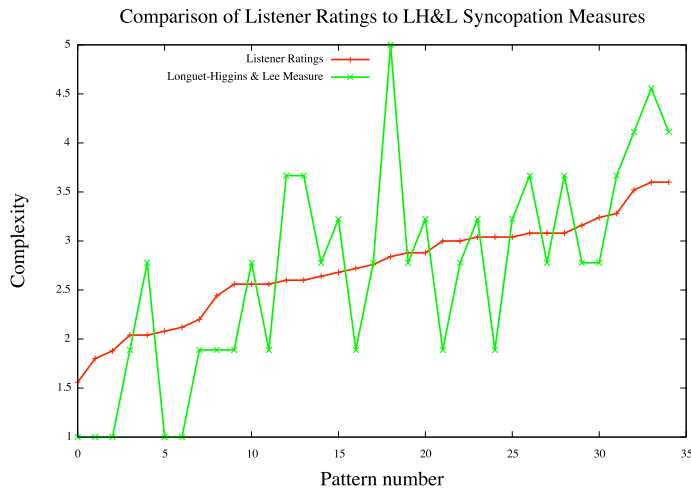Figure 3: Comparison of the LH&L model to listener ratings.

Figure 4: Comparison of the P&K metric salience variants to listener ratings.

using $S_{LH\&L}$ (although the difference is statistically insignificant). This could be due to the scaling of the P&K perceptual hierarchies reducing the magnitude of the syncopation measures such that there was not enough difference from the $S_{LH\&L}$ to produce significantly different predictions. Alternatively, the explanation may be that the lack of an overt metrical context in the listener ratings (Shmulevich and Povel, 2000) leads to disagreement with the judgements producing Palmer and Krumhansl's metric salience measures. The assumption that syncopations at different locations within the musical measure can be combined to a single value of syncopation may also be simplistic.

5 Conclusions

This preliminary work has implemented and tested LH&L's syncopation model against empirical data. This algorithm has been demonstrated to be applicable to alternative metrical salience parameterisation. This has allowed testing of P&K's perceptual hierarchy results and provides scope to test other alternative metrical salience measures and their generating models. Future work is to test the performance of the models against non-duple meter empirical data such as that reported by Essens (1995), and to incorporate the role of absolute tempo in the predictions (Handel, 1993).

References

Bharucha, J. J. (1993). MUSACT: A connectionist model of musical harmony. In S. M. Schwanauer and D. A. Levitt (Eds.), *Machine Models of Music*, pp. 497–510. Cambridge, Mass: MIT Press.
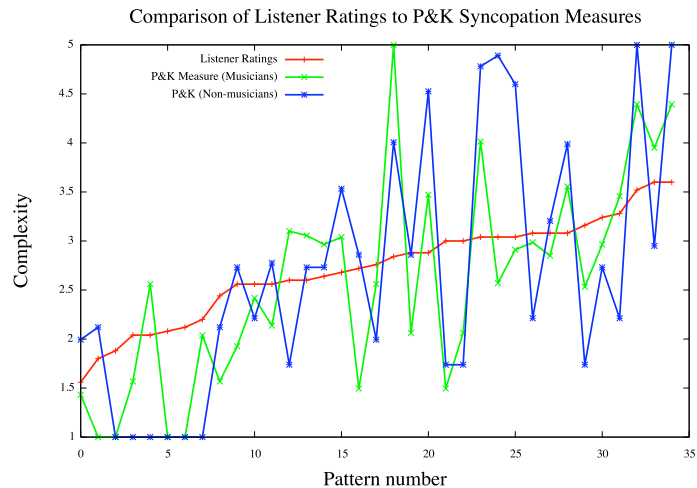
Bharucha, J. J. and P. M. Todd (1991). Modeling the perception of tonal structure with neural nets. In P. M. Todd and D. G. Loy (Eds.), *Music and Connectionism*, pp. 128–37. Cambridge, Mass: MIT Press.

Essens, P. (1995). Structuring temporal sequences: Comparison of models and factors of complexity. *Perception and Psychophysics 57*(4), 519–532.

Handel, S. (1993). The effect of tempo and tone duration on rhythmic discrimination. *Perception and Psychophysics 54*, 370–382.

Honing, H. (2006). Computational modeling of music cognition: A case study on model selection. *Music Perception 23*(5). in press.

Lerdahl, F. and R. Jackendoff (1983). *A Generative Theory of Tonal Music*. Cambridge, Mass: MIT Press. 368p.

Longuet-Higgins, H. C. and C. S. Lee (1984). The rhythmic interpretation of monophonic music. *Music Perception 1*(4), 424–41.

Meyer, L. B. (1956). *Emotion and Meaning in Music*. University of Chicago Press. 307p.

Palmer, C. and C. L. Krumhansl (1990). Mental representations for musical meter. *Journal of Experimental Psychology - Human Perception and Performance 16*(4), 728–41.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (2002). *Numerical Recipes in C++: The Art of Scientific Computing* (2nd ed.). Cambridge, U.K: Cambridge University Press.

Shmulevich, I. and D.-J. Povel (2000). Measures of temporal pattern complexity. *Journal of New Music Research 29*(1), 61–69.

Smith, L. M. (1999, June). *A Multiresolution Time-Frequency Analysis and Interpretation of Musical Rhythm*. Ph. D. thesis, Department of Computer Science, University of Western Australia.