Descriptors for Perception of Quality in Jazz Piano Improvisation

Jeff Gregorio, David S. Rosen, Michael Caro, and Youngmoo E. Kim
Drexel University
3401 Market Street
Philadelphia, PA
jgregorio@drexel.edu, drosen@drexel.edu, mcaro@drexel.edu, ykim@drexel.edu

Keywords

performance analysis, computational musicology and music analysis, jazz improvisation, expertise

ABSTRACT

Quality assessment of jazz improvisation is a multi-faceted, high-level cognitive task routinely performed by educators in university jazz programs and other discriminating music listeners. In this pilot study, we present a novel dataset of 88 MIDI jazz piano improvisations with ratings of creativity, technical proficiency, and aesthetic appeal provided by four jazz experts, and we detail the design of a feature set that can represent some of the rhythmic and harmonic attributes humans recognize as salient in assessment of performance quality. Inherent subjectivity in these assessments is inevitable, yet the recognition of performance attributes by which humans perceive quality has wide applicability to music information retrieval (MIR) community and jazz pedagogy. Preliminary results indicate that several musiciologically-informed features perform reasonably well in predicting performance quality labels via ordinary least squares regression.

1. INTRODUCTION

There is much literature describing the theory, techniques, and strategies which contribute to achieving mastery of jazz improvisation. The foundation of expert-level jazz improvisation relies on one's ability to deeply understand and creatively manipulate three critical aspects of a performance: rhythm, harmony, and melody. The potential variations in any of these areas is virtually infinite; however, jazz improvisation is a distinct type that includes the ability to generate the unforeseen, within the pre-existing structure of a song's chord structure, carefully balancing tradition and innovation [3]. Thus, while originality and creativity are essential to achieving high-quality improvisations, jazz's rich history and genre constraints must also be considered.

In this pilot study, we examine the efficacy of a set of computationally cheap, musicologically informed rhythmic and harmonic features in predicting expert listener perception of quality. Although the proposed features undoubtedly do not encompass every facet of improvisation, this work can potentially benefit musicians and music educators alike, providing some insight into how well-informed jazz listeners are affected by variations in rhythmic style and harmonic function relative to known chord changes.

We use an original symbolic (MIDI) dataset collected for a masters thesis by one of the authors (currently under review). This dataset is comprised of 88 improvisations from trained jazz pianists, with labels provided by jazz experts.

2. JAZZ IMPROVISATION DATASET

Jazz pianists ($N = 22$) were recruited from local university jazz programs, seminaries, and professional organizations in Philadelphia, PA. Trials were conducted at Drexel University in Philadelphia, PA. All performances took place in a professional sound booth using a 88-key semi-weighted MIDI controller keyboard, sustain pedal, music stand, and headphones. Apple's Logic Pro v.9.1.8 DAW was used to collect MIDI performance data, and provide a bass and drums audio accompaniment to a novel 16-bar chord sequence. Participants completed four takes, each consisting of four chord cycles (64 bars, $\approx$ 2 minutes).

To acquire ratings for the improvisations, four jazz experts were recruited as judges. These judges included a director of a collegiate jazz program, gigging professionals, and instructors, all with over 25 years of professional experience. Using the Consensual Assessment Technique (C.A.T.) [1], judges rated improvisations on a 7-point Likert scale for creativity, technical proficiency, and aesthetic appeal. All scales had excellent to very good reliability [4] using the intraclass correlation coefficient (ICC 2,1). These holistic ratings represent an evaluation of performative characteristics that contribute to the quality of an improvisation [2]. For each improvisation, ratings were averaged across categories and judges in order to arrive at a single quality of improvisation score ($M = 4.69$, $sd = .80$).

3. FEATURES

Feature design was guided primarily by common approaches to jazz improvisation in pedagogical literature [7] and cognitive studies [5] as well as questionnaires completed by the four judges regarding their rating criteria. The main focus of this pilot study is the development of time-series features computed on a sliding window over the duration of the performance. Descriptors used are statistics of each time series and its first difference, which include min, max, range, median, mean, standard deviation, skewness, and kurtosis.

3.1 Rhythmic Style

Multiple judges identified rhythm's primary importance, mentioning the importance of jazz's "swing" feel in assessing quality. Toward identifying perceptually relevant rhythmic patterns, we use an approach based on the inter-onset-interval (IOI) histogram called the Rhythmic Style Histogram Feature (RSHF) developed in [6], adapted to the MIDI modality. The IOI histogram is computed with a division of 6 bins per beat, sufficient to resolve quarter note triplets, eighth note triplets, and dotted sixteenth notes with a maximum IOI of 4 beats. We then compute each rhythmic style feature by computing the relative contribution of each bin to the total energy for patterns:

Feature	Note Durations
Duple	Whole, half, quarter, eighth, ...
Triple	Whole triplet, half triplet, quarter triplet, ...
2:1 Swing	Whole triplet + half triplet, half triplet + quarter triplet, ...
3:1 Swing	Dotted half + quarter, dotted quarter + eighth, ...

3.2 Harmonic Function Classes

We use various harmonic descriptors, each derived from the pitch class distribution of the performer's note choices. This distribution is computed on a sliding window over the duration of the performance. We use contextual information based on knowledge of the stimulus chord changes to group pitch classes by harmonic function relative to the current chord. The minimum chord length used in the stimulus is two beats, so we compute the pitch class distribution on a non-overlapping two-beat window. We use nine broad classes of harmonic function, as follows:

Class	Intervals/Pitch Classes (PCs)
Key Tones	PCs in the piece's key
Chord Tones	1st, 3rd, 5th, 7th scale degrees
Root Tones	1st and 5th scale degrees
Guide Tones	3rd and 7th scale degrees
Diatonic	PCs in chord's diatonic scale
Pentatonic	PCs in chord's pentatonic scale
'Avoid' Tones	3rds and 7ths that conflict with the chord's tonality (e.g. flat 3rd on a major 7th chord)
Color Tones	PCs in chord's diatonic scale, excluding chord tones (2nds, 4ths, and 6ths)
Dissonant Tones	Flat 2nds and sharp 4ths

3.3 Common Tone Voice Leading

The compositional and improvisational strategy of voice leading is generally defined as smooth motion between inner voices (notes) of a chord or melody through chord transitions [7]. One element of this strategy is known as *common tone* voice leading, whereby a composer or improviser will identify tones that are harmonically-related on either side of the transition. We make use again of harmonic function classes to look at the use of common-tone voice within each of the harmonic function classes.

The series of chords is parsed and the beat locations of chord transitions are identified. We then extract a two-beat window around each transition and get the relative contribution of the pitch class bins common to the same harmonic function class in both chords.

4. EXPERIMENT

The proposed feature set includes 336 statistical descriptors of the time series and their first differences. Such a high feature space dimensionality with many expected correlations necessitates dimensionality reduction prior to evaluating the predictive power of the features. We perform a two-step dimensionality reduction process via cross-validation, where each fold tests on four examples by one isolated performer, and trains on the remaining 84 examples.

The first round of cross-validation eliminates statistically insignificant features via Pearson's correlation coefficient, keeping only those significant at $p > 0.05$ in $2/3$ of all folds. This reduces the original 336 features to 163. Next, we use Principal Component Analysis (PCA) to eliminate redundancies by projecting the data into a space such that the variance of projected data is maximized and the basis dimensions are orthogonal. Through exhaustive cross-validation, we found projection into a single basis dimension yields both the highest adjusted R^2 (0.31) and lowest Mean Absolute Error (0.76). This dimension explains 53% of the dataset variance. The top contributors to dimension 1 are as follows, representing 75% of the feature weights:

Weight	Feature
0.87(39.1%)	hclass dissonant kurtosis
0.46(20.8%)	Δ hclass dissonant kurtosis
0.10(4.7%)	Δ rclass swing 2:1 kurtosis
0.09(4.2%)	hclass dissonant skewness
0.06(2.7%)	vlead guide kurtosis
0.06(2.5%)	rclass swing 2:1 kurtosis
0.03(1.3%)	Δ vlead guide kurtosis

This dimension correlates negatively with the quality labels ($\rho = -0.28, p = 0.01$). High kurtosis suggests large, sporadic deviations as opposed to smaller and more frequent ones. A possible interpretation is that low-scoring improvisers in the dataset used dissonance, swing, and common guide tone voice leading in an unfocused, inconsistent way.

5. CONCLUSIONS

From the MIDI data, we were able to compute a set of features that captured basic salient elements of rhythmic style, harmonic function, and common tone voice leading grounded in modern jazz theory and musicology. Though the proposed features addressed a small subset of qualities noted by judges as salient in their assessments, the four main time-series features contributing to the most informative PCA basis dimension were consistent with judge surveys indicating the importance of strategic use of dissonance, swing, and voice leading.

Though simple statistics time series and first difference features performed reasonably well, we note the shortcomings in using descriptors that solely capture the shape of the distribution. Future work will focus on the relationship between our existing features and specific cadences, as well as phrase-level descriptors based on melodic segmentation.

6. REFERENCES

[1] T. Amabile. The social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43:997–1013, 1982.

[2] R. Beaty, B. Smeekens, P. Silvia, D. Hodges, and M. Kane. A first look at the role of domain-general cognitive and creative abilities in jazz improvisation. *Psychomusicology: Music, Mind and Brain*, 23(4):262, 2013.

[3] P. F. Berliner. *Thinking in jazz*. University of Chicago Press, Chicago, Ill., 1994.

[4] J. Fleiss. *The Design and Analysis of Clinical Experimnts*. Wiley, New York, NY, 1986.

[5] P. N. Johnson-Laird. How jazz musicians improvise. *Music Perceptoin*, 19(3):415–442, 2002.

[6] M. Prockup, J. Scott, and Y. E. Kim. Representing musical patterns via the rhythmic style histogram feature. In *Proceedings of the ACM International Conference on Multimedia*, MM '14, pages 1057–1060, New York, NY, USA, 2014. ACM.

[7] R. Rawlins, N. Bahha, and B. Tagliarino. *Jazzology: The Encyclopedia of Jazz Theory for All Musicians*. Jazz Instruction Series. Hal Leonard, 2005.