Listeners' Perceived Emotions in Human vs. Synthetic Performance of Rhythmically Complex Musical Excerpts

Ève Poudrier[1], Bryan Jacob Bell[1], Jason Yin Hei Lee[2], and Craig Stuart Sapp[3]

[1] University of British Columbia, School of Music
[2]McGill University, Schulich School of Music
[3]Stanford University/CCARH/PHI
eve.poudrier@ubc.ca

Abstract. Research on listeners' perceived emotions in music draws on human and synthetic stimuli. Although research has shown that realistic synthetic audio can convey emotions, studies that compare listeners' experience of synthetic audio and human performances are limited. Using short musical excerpts, we investigate the effect of performance (human vs. synthetic) and instrumentation (piano vs. string quartet) as well as the influence of twelve musical features on participants' ratings of five emotional dimensions (mood, energy, movement, dissonance, and tension). Findings show a small main effect of performance and a large main effect of instrumentation. Synthetic audio was perceived as more positive in mood and less tense than human performances. Piano excerpts were also perceived as more positive and as conveying less tension and energy than synthetic excerpts. Several rhythmic and pitch measures were reliably predictive of participants' perceived emotions, supporting the need for considering finer-grain structural features when using naturalistic stimuli.

Keywords: empirical aesthetics, perceived emotion, computational musicology, music performance, synthetic audio generation

1 Introduction

Research on perceived emotion in music generally relies on listeners' judgments of aesthetic qualities based on audio excerpts of varying lengths. Such stimuli may involve pre-recorded human performances or synthetic audio generated by a computer following a set of instructions. Eerola and Vuoskoski (2013) report that a majority (75%) of studies in music and emotion research used human performances [1]. Although performance medium and source are usually reported along with the results, it is not clear whether the methods used to produce musical excerpts have an effect on listeners' experience. One disadvantage of using human performances as compared with synthetic audio generation is the lack of experimental control on the stimuli, which may limit researchers' ability to manipulate source materials and generalize findings.

1.1 Perceived Emotion in Human Performances versus Synthetic Audio

Research related to audio generation in terms of performance medium tends to focus on two aspects: timbral differences and expressive differences. Studies on timbral differences typically ask participants to identify and categorize single-note stimuli in terms of instrument type as well as perceptual dimensions such as "nasality," "brilliance," and "naturalness" [2]. Other studies investigate the effect of timbre on emotion by comparing excerpts played on different instruments (e.g., electronic synthesizer vs. human performances on piano, violin, and trumpet), with observed effects on listeners' perceived emotions interpreted as being related to acoustical factors [3], rather than performance medium.

Studies on expressive differences investigate expressive performance actions (micro-differences in terms of tempo, dynamics, articulation, intonation, and vibrato) applied in a human performance [4]. Most of the research in this area focuses on observed differences between the notated score and a human performance [5], or between different human performances of the same notated score [6]. Some studies have explored the effects of such differences on listeners' experience by manipulating human performances. For example, listeners have been shown to be able to distinguish between original and tempo-transformed versions of the same human performance [7]. Synthetic stimuli with different levels of timing manipulations have also been used to explore perceived "expressiveness" and "liveliness" [8], and the addition of other human-like expressive performance actions to synthetic audio, such as expressive dynamics, has been shown to result in higher ratings of "likeability" and "emotional expressiveness" [9]. Still, very few studies have explored listeners' ratings of emotional expression in synthetic audio as compared to human performances. On one hand, listeners have been found to have a negative bias in their ratings of expressive qualities of human performances presented as synthetic (i.e., "pseudo-synthetic" performances) [8]. On the other hand, synthetic versions of short melodies with human-like expressive differences in tempo, sound level, spectrum, articulation, attack, vibrato, and timing has been shown to convey discrete emotions such as happiness, sadness, anger, fear, and tenderness as effectively as human performances of the same melodies [10]. Nevertheless, direct comparisons of listeners' perceived emotion in human and synthetic performances of multi-part music are still needed.

In this study, we use rhythmically complex musical excerpts characterized by concurrent rhythmic patterns that cannot readily be mapped onto a single metric grid (i.e., "polyrhythm"). These materials were selected because they provide a naturalistic and rich environment within which listeners' perceived emotions can be tackled. To-date, very little research has looked into how rhythmically complex music is aesthetically evaluated by listeners. When real music is used (as opposed to controlled "lab" stimuli), attention is devoted to global aspects of the musical compositions, such as tempo, loudness, timbre, and mode, among other factors. However, most studies do not offer sufficient fine-grained information on the rhythmic structure of the selected music to allow for generalization over a wider range of music.

1.2 Aims

The goal of the present study is twofold. First, we aim to determine whether performance (human vs. synthetic) has an effect on listeners' judgment of five emotional dimensions (mood, energy, movement, dissonance, and tension) for two different instrumentation types (piano and string quartet). Second, because we used excerpts from musical compositions that feature complex rhythmic and harmonic structures, we also aim to explore the effect of features specific to the musical style on listeners' perceived emotions. To that end, we used a set of computational measures of rhythmic structure (duration, density, alignment, contrast, and regularity) as well as pitch organization (pitch range, pitch mean, register, and sonority dissonance).

We assumed a null hypothesis for the influence of performance, but predicted a main effect of instrumentation. With regard to musical features, we expected event density to be positively correlated with mood, energy, and tension [11, 12, 13], pitch range to be positively correlated with mood and energy, and pitch mean to be negatively correlated with energy [14]. We also expected sonority dissonance to be predictive of perceived dissonance and tension.

2 Methods

2.1 Participants

Participants were recruited using an online survey implemented in Qualtrics. The survey was approved by the Ethics Review Board of the University of British Columbia, and shared through social media postings, email notifications to institutional and professional listservs, and the UBC Psychology SONA platform. 162 participants with normal hearing completed the study; two datasets were excluded from analysis due to reported difficulty with English in everyday life. Gender distribution was uneven, with a large proportion of participants self-identifying as women (76%) as compared to men (21%); two participants self-identified as non-binary persons and three participants selected "prefer not to answer." Participants' age ranged from 18 to 59 years old (M = 23.3; SD = 7.3), and self-reported years of formal musical training ranged from 0 to 20 years (M = 5.9; SD = 5.1). A greater proportion of participants reported familiarity with the musical style represented by the excerpts (43%, as compared with 23% and 34% for no familiarity and "not sure"), but much fewer participants reported familiarity through listening or performance of a specific excerpt (23%, as compared with 61% and 17% for no familiarity and "not sure"). Finally, most participants listened to the excerpts using built-in speakers (41%), followed by standard and noise-canceling headphones or earbuds (26% and 25%); a small proportion of participants reported using external speakers (8%), while only one participant reported using a phone speaker.

2.2 Materials

Sixteen musical excerpts from 12 different composers were selected from the *Suter (1980) Corpus* [15], ranging from 1893 to 1965 in terms of composition year (see Table 1).[1] Based on the availability of realistic audio synthesis and for contrast in timbre, we

[1] A full list of examples from the *Suter (1980) Corpus* and associated metadata is available at: https://polyrhythm.humdrum.org. The examples used in this experiment are available in kern format at: https://github.com/polyrhythm-project/rds-scores/tree/master/experiment-lmf1.

selected an equal number of short piano and string quartet examples with a duration of 5 to 9 s (M = 7.2; SD = 1.1).

Table 1. Source musical compositions for experimental stimuli listed alphabetically by composer's last name. There are eight examples for each instrumentation type.

Composer	Work Title	Instrumentation	Year
Bartók, Béla	Romanian Folk Dances	Piano	1915
	Piano Sonata	Piano	1926
	String Quartet No. 3	String Quartet	1927
Berg, Alban	Lyric Suite	String Quartet	1926
Britten, Benjamin	String Quartet No. 2, op. 36	String Quartet	1945
Debussy, Claude	String Quartet, op. 10	String Quartet	1893
Falla, Manuel de	"Jota", from Seven Spanish Songs	Piano	1914
Gershwin, George	Rhapsody in Blue	Piano	1924
Hindemith, Paul	String Quartet, op. 10	String Quartet	1918
Ives, Charles	String Quartet No. 1	String Quartet	1909
Martin, Frank	Prelude No. 8	Piano	1948
	Esquisse	Piano	1965
Martinů, Bohuslav	String Quartet No. 7	String Quartet	1947
Prokofiev, Sergei	Piano Sonata No. 7, op. 83	Piano	1942
	Piano Sonata No. 9, op. 103	Piano	1947
Ravel, Maurice	String Quartet	String Quartet	1903

Two audio versions of each example were prepared: a human performance extracted from a commercial recording randomly selected from available recordings in the Naxos Music Library, and a high-quality musical instrument digital interface (MIDI) rendering using the EastWest sound library. Audio files extracted from recorded examples were trimmed using Audacity to allow excerpts' duration to be more precisely measured. Synthetic examples were fine-tuned in terms of MIDI note velocity (i.e., volume of each note) and articulations (legato vs. staccato for piano, but a wide variety of options for strings) to match the human performances as closely as possible. The precise timing of raising and lowering the sustain pedal was also fine-tuned for piano excerpts. The tempo of synthetic examples was set to match the average tempo of the human recorded performances, but without rubato or expressive microtiming (i.e., the timing of individual note onsets or releases). To match the acoustics of the human recordings as closely as possible, reverb was added to the piano examples using Logic's Space Designer; it was not deemed necessary to add reverb to the string quartet examples, which fairly closely matched the acoustics of the human performances. A 0.2 s fade-out was applied to the end of each example to reduce the abruptness of the ending, and both audio file versions were then amplified or attenuated to a peak volume of -1 dB.

2.3 Procedure

The experiment was conducted online using Qualtrics, with participants instructed to complete the tasks in one sitting, focusing only on doing the experiment, and in a quiet location or wearing noise-canceling earphones. The order of the experimental trials was randomized across musical excerpts so that each participant heard one performance version (human or synthetic) of each of the sixteen excerpts. To avoid bias, participants were not informed of the type of performance they would hear. Participants were instructed to listen to the excerpt in its entirety at least once, and then rate the excerpt using five seven-point Likert scales. Participants rated the perceived mood (negative–positive), energy (low–high), movement (very little–very much), dissonance (low–high), and tension (low–high), with "movement" referring to how much the participant felt that they could move along to the music.

First, participants provided consent, and read the survey instructions. Participants completed a pre-experiment questionnaire on which they reported their age, gender, formal musical training, and English-language fluency. Prior to listening to the experimental stimuli, participants heard a short audio file during which they were instructed to adjust their volume to a comfortable level, and then completed two practice trials (one of each instrumentation type). At the end of the survey, participants were asked to report what listening device was used to complete the survey, and whether they were familiar with the musical style of the excerpts or with the excerpts themselves through listening or performance. Lastly, participants were given the opportunity to provide feedback and read a debriefing document.

2.4 Measures

In addition to participants' ratings of the five dependent variables using seven-point Likert scales, we selected several measures derived from rhythmic and pitch structures to explore the relationship between musical features and participants' perceived emotions. Rhythmic features required visualization and analysis of each excerpt and assessment of the differences between concurrent rhythmic patterns. Instrumental parts were divided into two contrasting rhythmic groups (A and B) based on rhythmic similarity within the group and dissimilarity across groups, with the lowest part on the score assigned to Group A by default. The experimental excerpts include up to four instrumental parts; note that although piano excerpts are notated on two staves, each staff could include more than one part. Because the rhythmic design of a given instrumental part may vary over time, group attribution was performed at the measure level. To allow for comparison between examples with a different number of instrumental parts, we use composite rhythms, i.e., the sequential presentation of event onsets across parts. Figure 1 illustrates the visual analytic markup for a sample used in the experiment. Group A notes are colored in red, while Group B notes are in blue. The top two staves are the original score and underneath are the extracted rhythmic patterns and number of event onsets for Group A only, Group B only, Groups A and B combined ("composite"), and the intersection of Groups A and B ("coincidence"). The analytic markup shown in the musical example is automatically generated by the *composite* filter in Verovio Humdrum Viewer [16]; full documentation for the composite filter is available at: https://doc.verovio.humdrum.org/filter/composite.

Fig. 1. Visual analytic markup for Gershwin, Rhapsody in Blue (1924), mm. 91–94.

Table 2 presents the six features used to characterize rhythmic structure. Four additional pitch features were also used. *Pitch mean* (average pitch) and *pitch range* (interval between lowest and highest pitch) are calculated using MIDI note values. *Register* corresponds to the proportion of events in each of three ranges: *low* (below C3), *middle* (C3 to C5), and *high* (above C5). To measure *sonority dissonance*, each sonority was assigned a score based on its most dissonant interval (octave/unison = 0; P4/P5 = 1; M/m 3/6 = 2; M2/m7/M9 = 3; A4/d5/m9 = 4; m2 = 5); these values were then averaged and weighed by duration.

Table 2. Calculation and interpretation of rhythmic features

Feature	Calculation	Interpretation
Duration	Total duration of audio file in seconds	N/A
Composite event density	Total number of composite events divided by audio file duration	Rate of presentation of events in time (global information load)
Event density ratio	Number of events in smaller group divided by number of events in larger group	Potential for metric ambiguity or conflict across groups
nPVI group difference[2]	Absolute difference between the nPVI scores of the two rhythmic groups	Contrast in note-to-note regularity across rhythmic groups
Nested ratio	Total number of coinciding event onsets across rhythmic groups divided by total number of composite events	Potential for integrated percept of two rhythmic groups
Polarity ratio	Absolute difference between number of events in rhythmic groups divided by total number of composite events	Relative activity within rhythmic groups (salience)

[2] This measure is an extension of the *normalized pairwise variability index*, a measure of the average durational variation between successive pairs of events.

2.5 Analysis

We conducted statistical analysis in RStudio, with R version 4.1.1 and used the rstatix package for summary statistics and the broom package for summarizing linear models.[3] Although piano and string quartet examples were different in terms of musical materials, they belong to the same historical period. The relative stylistic homogeneity of these musical excerpts was supported by a series of *t* tests: there was no statistically significant difference between piano and string quartet for each of the twelve musical features. To test the effect of performance and instrumentation on participant ratings, we conducted a two-way Multivariate Analysis of Variance (MANOVA) on the combined five dependent variables with performance and instrumentation as the independent variables. Point biserial correlations were used to explore the linearity between the five dependent variables and the two independent variables of performance and instrumentation. To explore the effects of our twelve musical features on participants' ratings, we performed multiple regression analyses. A linear model was constructed between each dependent variable and the twelve musical features. Because participants' ratings were done on a seven-point Likert scale, dependent variables were log-transformed using $\log(1+x)$.

3 Results

Participants rated 16 excerpts on five Likert scales ($N = 2{,}560$). The average rating for energy was the highest ($M = 4.8$; $SD = 1.5$), while those for dissonance ($M = 3.5$; $SD = 1.7$) and movement ($M = 3.7$; $SD = 1.8$) were the lowest. The average ratings for mood and tension were in the 4–5 range ($M = 4.3$ and 4.0; $SD = 1.5$ and 1.7).

3.1 Performance and Instrumentation

The main effect of performance was statistically significant, but small, $F(1, 2556) = 2.89$, $p = .013$, while the main effect of instrumentation was statistically significant and large, $F(1, 2556) = 15.97$, $p < .001$. There was also a significant, although relatively small, interaction between performance and instrumentation, $F(1, 2556) = 3.32$, $p = .005$.

Point biserial correlations were calculated between each dependent variable and performance (Human = 1; Synthetic = 2) as well as instrumentation (Piano = 1; String Quartet = 2). Performance was positively correlated with mood, $r_{pb}(2558) = .05$, $p = .01$, but negatively correlated with tension, $r_{pb}(2558) = -.06$, $p = .002$. Instrumentation was negatively correlated with mood, $r_{pb}(2558) = -.09$, $p < .0001$, but positively correlated with energy and tension, $r_{pb}(2558) = .12$, $p < .0001$, and $r_{pb}(2558) = .13$, $p < .0001$. In other words, participants perceived synthetic excerpts as more positive in mood and as conveying less tension than human performances. Piano excerpts were also perceived as more positive in mood as well as lower in energy and tension than string quartet excerpts.

[3] RStudio, rstatix, and broom are available at: https://www.R-project.org/, https://CRAN.R-project.org/package=rstatix, and https://CRAN.R-project.org/package=broom.

3.2 Effects of Musical Features

A summary of the parameter estimates for each of the five dependent variables and the twelve musical features is presented in Table 3 (rhythmic features) and Table 4 (pitch features). All twelve musical features were predictive of participants' ratings for one or more emotional dimensions, with the best model accounting for more than a third of the variance in participants' ratings of energy (R^2 = .398).

Rhythmic Features. Event density ratio, nested ratio, and polarity ratio were the most reliable predictors for four of the five emotional dimensions with significance levels of $p < .001$. Event density ratio and polarity ratio were negatively correlated with mood and movement, and positively correlated with tension. Excerpts with a higher potential for metric ambiguity or conflict and greater contrast in the number of events within each rhythmic group were perceived as less positive in mood, less likely to induce movement, and as conveying more tension. On the other hand, although both factors were also predictive of participants' ratings of dissonance, higher event density ratio was predictive of higher perceived dissonance, while higher polarity ratio was predictive of lower perceived dissonance. In contrast, nested ratio was positively correlated with mood, movement, and dissonance, but negatively correlated with tension. Excerpts that featured more coinciding events were perceived as more positive, more likely to induce movement, more dissonant, but less tense. nPVI group difference was predictive of participants' ratings for three of the five emotional dimensions. A greater contrast between groups in note-to-note rhythmic regularity was correlated with a more positive mood, higher energy, and lower perceived tension. On the other hand, duration and composite density had a relatively limited effect on participants' ratings. Excerpts' duration was negatively correlated with mood and positively correlated with tension. Longer excerpts were perceived as less positive in mood and as conveying more tension. Composite event density was predictive of participants' ratings for energy, with higher composite density predictive of higher energy ratings.

Pitch Features. The influence of pitch-related features on participants' ratings of perceived emotions was small, but not negligible. Pitch mean was negatively correlated with energy and dissonance, with higher pitch mean being predictive of lower perceived energy and dissonance. Pitch range was also reliably predictive of participants' ratings for mood, energy, and movement with significance levels of $p < .001$. Larger range was correlated with a more positive mood, higher energy level, and a greater impulse to move along with the music. Register (low, middle, and high) was predictive of participants' ratings of mood, energy, and tension with significance levels of $p < .001$. A larger proportion of events in any one of the three registers was positively correlated with mood, but negatively correlated with energy and tension. In other words, the concentration of events in one register, rather than a specific register or a more balanced dispersion of pitch activity, was perceived as more positive in mood, but as conveying lower energy and less tension. As expected, sonority dissonance was positively correlated with perceived dissonance, but the correlation with tension was small and not significant.

Table 3. Parameter estimates for rhythmic features and each dependent variable. Significance levels are as follows: '***' $p < .001$; '**' $p < .01$; '*' $p < .05$.

Dependent Variable	Parameter	Estimate	Std. Error	*t* value	Pr(>|*t*|)
Mood	(Intercept)	-417.70	39.85	-10.48	< .001***
($R^2 = 0.282$)	Duration	-0.14	0.01	-9.72	< .001***
	Composite event density	0.01	0.00	1.67	.10
	Event density ratio	-5.62	0.61	-9.17	< .001***
	nPVI group difference	0.01	0.00	5.03	< .001***
	Nested ratio	1.32	0.14	9.59	< .001***
	Polarity ratio	-5.09	0.54	-9.48	< .001***
Energy	(Intercept)	212.80	34.97	6.09	< .001***
($R^2 = .398$)	Duration	-0.01	0.01	-0.53	.60
	Composite event density	0.03	0.00	7.08	< .001***
	Event density ratio	0.71	0.54	1.32	.19
	nPVI group difference	0.00	0.00	-3.23	.001**
	Nested ratio	-0.06	0.12	-0.49	.62
	Polarity ratio	0.50	0.47	1.07	.29
Movement	(Intercept)	-107.50	59.39	-1.81	.07
($R^2 = .107$)	Duration	-0.04	0.02	-1.94	.05
	Composite event density	0.00	0.01	0.47	.64
	Event density ratio	-4.16	0.91	-4.55	< .001***
	nPVI group difference	0.00	0.00	0.47	.64
	Nested ratio	1.24	0.20	6.07	< .001***
	Polarity ratio	-4.19	0.80	-5.24	< .001***
Dissonance	(Intercept)	105.50	55.96	1.89	.06
($R^2 = .113$)	Duration	0.04	0.02	1.68	.09
	Composite event density	0.00	0.01	-0.15	.88
	Event density ratio	4.22	0.86	4.90	< .001***
	nPVI group difference	0.00	0.00	-0.19	.85
	Nested ratio	-1.32	0.19	-6.87	< .001***
	Polarity ratio	4.14	0.75	5.49	< .001***
Tension	(Intercept)	557.70	49.97	11.16	< .001***
($R^2 = .260$)	Duration	0.14	0.02	7.28	< .001***
	Composite event density	0.00	0.01	0.59	.56
	Event density ratio	5.00	0.77	6.51	< .001***
	nPVI group difference	-0.01	0.00	-4.47	< .001***
	Nested ratio	-1.35	0.17	-7.84	< .001***
	Polarity ratio	4.40	0.67	6.54	< .001***

Table 4. Parameter estimates for pitch features and each dependent variable. Significance levels are as follows: '***' $p < .001$; '**' $p < .01$; '*' $p < .05$; intercept and R^2 values are the same as in Table 3.

Dependent Variable	Parameter	Estimate	Std. Error	*t* value	Pr(>|*t*|)
Mood	(Intercept)	-417.70	39.85	-10.48	< .001***
(R^2 = .282)	Pitch mean	0.00	0.01	-0.34	.73
	Pitch range	0.01	0.00	8.30	< .001***
	Register low	423.90	40.05	10.58	< .001***
	Register middle	424.90	40.14	10.59	< .001***
	Register high	424.60	40.22	10.56	< .001***
	Sonority dissonance	0.00	0.02	-0.01	.99
Energy	(Intercept)	212.80	34.97	6.09	< .001***
(R^2 = .398)	Pitch mean	-0.03	0.00	-5.87	< .001***
	Pitch range	0.01	0.00	11.64	< .001***
	Register low	-211.00	35.14	-6.00	< .001***
	Register middle	-210.20	35.22	-5.97	< .001***
	Register high	-210.60	35.29	-5.97	< .001***
	Sonority dissonance	-0.04	0.02	-1.79	.07
Movement	(Intercept)	-107.50	59.39	-1.81	.07
(R^2 = .107)	Pitch mean	-0.01	0.01	-1.36	.18
	Pitch range	0.01	0.00	7.55	< .001***
	Register low	112.70	59.69	1.89	.06
	Register middle	113.70	59.81	1.90	.06
	Register high	112.80	59.94	1.88	.06
	Sonority dissonance	-0.12	0.04	-3.38	< .001***
Dissonance	(Intercept)	105.50	55.96	1.89	.06
(R^2 = .113)	Pitch mean	-0.02	0.01	-2.32	.02*
	Pitch range	0.00	0.00	-1.52	.13
	Register low	-107.50	56.24	-1.91	.06
	Register middle	-107.30	56.36	-1.90	.06
	Register high	-106.50	56.47	-1.89	.06
	Sonority dissonance	0.15	0.03	4.20	< .001***
Tension	(Intercept)	557.70	49.97	11.16	< .001***
(R^2 = .260)	Pitch mean	0.00	0.01	-0.02	.98
	Pitch range	0.00	0.00	-1.23	.22
	Register low	-560.30	50.23	-11.15	< .001***
	Register middle	-561.00	50.33	-11.15	< .001***
	Register high	-561.50	50.44	-11.13	< .001***
	Sonority dissonance	-0.01	0.03	-0.46	.65

4 Discussion

The first goal of our study was to investigate more directly the influence of synthetic generation, as compared to, human performance on listeners' perceived emotions in rhythmically complex music excerpts that contrasted in acoustics (piano and string quartet). Performance medium was found to have a small but significant effect, with synthetic performances being perceived as more positive in mood and as conveying less tension. This finding extends previous research that showed that synthetic generation of short melodies can convey discrete emotions effectively [10], and further qualifies the effect of synthetic audio on listeners' perceived emotions. As expected, instrumentation also had a significant and large effect, with piano excerpts giving rise to higher valence and arousal judgments as well as lower ratings for tension. The main effect of instrumentation is consistent with research on the influence of timbre and the influence of acoustical factors on listeners' perceived emotions [3]. While musical excerpts varied across piano and string quartet, they were very similar in terms of the specific musical features considered. Nonetheless, the presence of some hidden factor related to musical excerpts cannot be fully discounted and should be taken into consideration in future experiments (i.e., using musical examples that afford both piano and string quartet performances).

Our second goal was to explore the influence of a number of rhythmic and pitch features on listeners' perceived emotions. Many of our findings are novel and open avenues of investigation on the role of rhythmic structure on perceived emotion. Most notably, event density ratio, a measure of the probability of metric ambiguity or conflict between parts, and polarity ratio, a measure of the contrast in the number of events across rhythmic groups, were predictive of perceived mood, movement, and tension. Similarly, the degree of coinciding event onsets between rhythmic layers (i.e., nested ratio) had reliable, but contrasting effects on participants' ratings of mood, movement, dissonance, and tension. Taken together, these results suggest that rhythmically more integrated musical parts are perceived as more positive and are more likely to induce movement. This is consistent with findings that higher levels of rhythmic complexity have a negative effect on entrainment, which may result in reduced enjoyment [13]. Our results also point to an interaction between rhythmic structure and perceived dissonance by which a lower degree of integration of concurrent rhythmic streams may reduce listener's sensitivity to sonority dissonance between parts. To our knowledge, this is a yet unexplored area that warrants further investigation.

Pitch features had a smaller, but not negligible effect on listeners' ratings. In addition to the expected links between pitch range and higher mood and energy as well as between pitch mean and lower arousal, a wider pitch range was found to be correlated with induced movement. Also of note is the observed relationship between the concentration of pitch in one register, rather than one specific register, being predictive of a more positive mood as well as lower levels of perceived energy and dissonance. Both of these findings warrant further study using a wider selection of music and more controlled stimuli.

Several of our findings show that the interaction of rhythmic patterns in multi-part music plays a significant role in listeners' emotional experience, which calls into question the ecological validity of studies that focus on relatively simple musical sequences to study perceived emotion. Overall, findings based on rhythmic and pitch features suggest that more attention should be devoted to finer-grain musical features and their potential effect on listeners' emotional experience.

5 Acknowledgments

The authors thank the members of the UBC Rhythm Computation and Cognition Lab for the encoding of the musical excerpts and Walker Williams for audio preparation. This research is supported in part by funding from the Social Sciences and Humanities Research Council of Canada.

6 References

1. Eerola, T., Vuoskoski, J. K.: A review of music and emotion studies: Approaches, emotion models, and stimuli. Music Perception: An Interdisciplinary Journal 30(3), 18–49 (2013).
2. Kendall, R. A., Carterette, E. C., Hajda, J. M.: Perceptual and acoustical features of natural and synthetic orchestral instrument tones. Music Perception: An Interdisciplinary Journal 16(3), 327–363 (1999).
3. Hailstone, J. C., Omar, R., Henley, S. M. D., Frost, C., Kenward, M. G., Warren, J. D.: It's not what you play, it's how you play it: Timbre affects perception of emotion in music. Quarterly Journal of Experimental Psychology 62(11), 2141–2155 (2009).
4. Kirke, A., Miranda, E.: A survey of computer systems for expressive music performance. ACM Computing Surveys 42(1), Article 3 (2009).
5. Oore, S., Simon, I., Dielemen, S., Eck, D. Simonyan, K.: The time with feeling: Learning expressive musical performance. Neural Computing and Applications 32, 955–967 (2020).
6. Repp, B. H.: A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. Journal of the Acoustical Society of America 104(2), 1085–1100 (1998).
7. Honing, H.: Is expressive timing relational invariant under tempo transformation? Psychology of Music 35(2), 276–285 (2007).
8. Hähnel, T., Berndt, A.: Synthetic and pseudo-synthetic music performances: An evaluation. In: Proceedings of the 3rd International Conference of Students of Systematic Musicology (SysMus10), Cambridge, United Kingdom (2010).
9. Kamenetsky, S. B., Hill, D. S., Trehub, S. E.: Effect of tempo and dynamics on the perception of emotion in music. Psychology of Music 25(2), 149–160 (1997).
10. Juslin, P. N.: Perceived emotional expression in synthesized performances of a short melody: Capturing the listener's judgment policy. Musicae Scientiae 1(2), 225–256 (1997).
11. Gabrielson, A., Juslin, P. N.: Emotional expression in music performance: Between the performer's intention and the listener's experience. Psychology of Music 24(1), 68–91 (1996).
12. Fernández-Sotos, A., Fernández-Caballero, A., Latorre, J. M.: Influence of tempo and rhythmic unity in musical emotion regulation. Frontiers in Computational Neuroscience 10, Article 80 (2016).
13. Labbé, C., Grandjean, D.: Musical emotions predicted by feelings of entrainment. Music Perception: An Interdisciplinary Journal 32(2), 170–185 (2014).
14. Gomez, P., Danuser, B.: Relationships between musical structure and psychophysiological measures of emotion. Emotion 7(2), 377–387 (2007).
15. Poudrier, È., Shanahan, D.: Modeling rhythmic complexity in a corpus of polyrhythm examples from Europe and America, 1900–1950. In: Proceedings of ICMPC15/ESCOM10, ed. R. Parncutt and S. Sattman, pp. 355–360. Centre for Systematic Musicology, University of Graz, Austria (2018).
16. Poudrier, È., Sapp, C. S.: Polyrhythm analysis using the *composite* tool. In: 9th International Conference on Digital Libraries for Musicology (DLfM2022), July 28, 2022, Prague, Czech Republic. ACM, New York, NY, United States (2022).