THE SIGNIFICANCE OF THE LOW COMPLEXITY DIMENSION IN MUSIC SIMILARITY JUDGEMENTS

Jeff Ens
Simon Fraser University
jeffe@sfu.ca

Bernhard E. Riecke
Simon Fraser University
ber1@sfu.ca

Philippe Pasquier
Simon Fraser University
pasquier@sfu.ca

ABSTRACT

Previous research has demonstrated that similarity judgements are context specific, as they are shaped by cultural exposure, familiarity, and the musical aesthetic of the content being compared. Although such research suggests that the criterion for similarity judgement varies with respect to the musical style of the content being compared, the specific musical factors which shape this criterion are unknown. Since dimensional complexity differentiates musical genres, and has been shown to affect similarity judgements following lifelong exposure, this experiment investigates the short-term influence of dimensional complexity on similarity judgements. Rhythmic and pitch sequences with two levels of complexity were factorially combined to create four distinct types of prototype melodies. 51 participants rated the similarity of each type of prototype melody (M) to two variations, one in which the pitch content was modified $(\bar{M}_p)$, and another in which the rhythmic content was modified $(\bar{M}_r)$. The results indicate that rhythm and pitch complexity both play a significant role, influencing the perceived similarity of $\bar{M}_p$, and $\bar{M}_r$. The dimension bearing low complexity information was found to be the predominant factor in similarity judgements, as participants found modifications to this dimension to significantly decrease perceived similarity.

1. INTRODUCTION

Similarity directly informs our experience of music, enabling the perception of cohesion within a musical work, and the categorization of musical works. Consequently, developing models that encapsulate the manner in which similarity is perceived, is of critical importance within the areas of Musicology, Music Cognition and Music Theory [30]. In particular, the search for robust and flexible similarity measures has dominated research in the Music Information Retrieval (MIR) domain, as large digital databases of music information necessitate content-based querying and retrieval, and classification. Although there is a large body of research that explores similarity perception within music, many aspects of similarity perception are not yet fully understood. The current study corroborates previous evidence that similarity criterion vary with respect to the musical content being compared [9], demonstrating that the complexity of pitch and rhythmic content influence similarity perception.

Since pitch and rhythm are the two most prominent musical dimensions in the context of symbolic notation, the current study will manipulate complexity along these dimensions and observe the effects on similarity perception. Although no musical dimensions are completely orthogonal, as a modification in a particular dimension may affect the perception of other dimensions, the complexity of pitch and rhythmic content can be measured independently, and there is evidence that these dimensions are processed separately in cognition [13, 27]. Therefore, pitch and rhythm complexity were considered to be independent for the purposes of this study. *Pitch content* refers to the sequence of pitches encapsulated in a particular melody, and *rhythm content* refers to the sequence of durations. *Dimensional complexity* refers to the absolute level of complexity along a particular musical dimension. In this study we measure the dimensional complexity of pitch and rhythm content.

2. RELATED WORK

Previous work examining the perception of musical similarity, has focused on establishing a hierarchy of musical dimensions, ranking their observed contributions to similarity perception. On a whole, most research claims that rhythmic information is the most important. Halpern [7] constructed 16 melodies — a factorial combination of two pitch sequences, two rhythmic sequences, two tonal structures and forward and reversed versions — and found that rhythm was the most important distinguishing factor, followed by pitch, direction and tonal structure. Similarly, Rosner and Meyer [19] found rhythm to be the strongest determinant of melodic similarity. Despite the general consensus that rhythm plays a dominant role in similarity judgements, pitch still plays a considerable role. Dowling [2] demonstrated that a modified imitation of a prototype melody is often misidentified as the prototype when it has a similar pitch contour.

Given the multidimensional nature of music, many researchers have found it useful to make the distinction between surface-level and structural features. In gen-

 Attribution: Jeff Ens, Bernhard E. Riecke, Philippe Pasquier. "The Significance of the Low Complexity Dimension in Music Similarity Judgements", 18th International Society for Music Information Retrieval Conference, Suzhou, China, 2017.

eral, surface-level attributes include contour, loudness and tempo while structural attributes denote aspects of form, thematic development and patterns. In short term contexts, where participants are unfamiliar with the musical material being compared, surface-level features are a strong predictor of both melodic [15, 19, 22] and polyphonic [9] similarity. Prince [15] found that rhythm was the dominant aspect informing perceived melodic similarity, followed by contour, meter, and tonal structure.

However, there is increasing evidence which questions the generality of these results, as contextual factors including familiarity, cultural exposure, and the aesthetic of the musical content being compared, have been shown to have a considerable effect on similarity perception. Pollard-Gott found that with repeated listening, surface level features became less influential and thematic material became more important [14]. Similarly, the long term analysis of a collection of folk melodies by a panel of experts, placed emphasis on thematic and motivic similarity above all other factors [31]. Schubert and Stevens [22] found that contour is more important than harmonic structure for making similarity comparisons, but with musical expertise, harmonic structure also has an effect.

Other research has shown that cultural exposure affects similarity perception. Hannon and Trehub [8] found the metrical bias of North American adults to be the result of an enculturation processes, with no evidence of a natural predisposition for the simple meters which characterize much of western music. Goldstone [6] suggests that humans learn by focusing on perceptual features that are more informative, at the cost of decreased attention towards other dimensions. This phenomenon has been observed in a musical context, where the voice that consists of immediate and exact repetitions of a short musical fragment tends to perceptually decrease in salience for the listener over time [24]. Instead, the listener is naturally drawn to focus on the high complexity voice. Since distinct rhythmic durations occur at a relatively higher frequency than distinct pitches in western music, they demand less attention than pitch content. After years of exposure, this likely results in an increased sensitivity to the pitch content in a melody [17]. Notably, Eerola et al. [3] demonstrated that musical complexity perceptions are shaped by exposure to different musical culture, which likely results from the mechanisms described above.

In addition to the factors mentioned above, music aesthetic has been shown to influence how similarity is perceived. Lamont and Dibben [9] examined similarity relationships in two contrasting musical styles, requiring participants to rate the similarity of extracts from a Beethoven sonata (op. 10, no. 1, first movement) and a dodecaphonic work composed by Schoenberg (Klavierstück op. 33a). Nine polyphonic excerpts were selected from each piece, each approximately eight measures long, and the similarity of each possible combination was rated by participants, resulting in 36 similarity ratings for each piece. Notably, both pieces are composed for solo piano, and have more than one theme which is developed throughout the duration of each work. They found that similarity judgements were primarily based on surface level features, however, the similarity judgements for each piece were predominantly influenced by different surface features. These results suggested that each piece establishes a different similarity criterion within which listeners make appropriate similarity judgements. Although Lamont and Dibben demonstrated that the criterion for similarity judgements varies with respect to the musical aesthetic of the stimuli being compared, the specific musical factors which caused this phenomenon are still unknown, directly motivating our experiment.

3. MOTIVATION

As evidenced by the brief overview in section 2, numerous studies have demonstrated the prevalent influence of contextual factors on musical similarity judgements [8, 9, 14, 17, 31], directly motivating further study in this area. Since contextual factors like cultural exposure and familiarity are difficult to integrate into a similarity measure, this study examines the third contextual factor, the role of the musical content itself in shaping a criterion for similarity judgements. The phenomenon that Lamont and Dibben [9] observed, provides evidence that musical content influences the manner in which music is compared, as participants used different musical dimensions to make comparisons depending on the nature of the musical content. In light of this evidence, it is worthwhile to examine how specific musical characteristics of the content being compared shape similarity judgements, which does not appear to have been examined previously. Due to the fact that dimensional complexity differentiates musical genres [3], and affects similarity judgements following lifelong exposure [8], this experiment investigates the short-term influence of dimensional complexity on melodic similarity judgements. More specifically, this study investigates the role of dimensional complexity in shaping awareness to modifications in that particular dimension, effectively establishing a criterion for melodic similarity judgements.

Previous research has shown that limitations on the human capacity for musical memory, have an effect on musical perception. Participants found it more difficult to retain melodies with complex contours, which were devoid of any repetition, and were often unable to distinguish them from another complex contour [18]. Moreover, complexity was one of four variables which collectively predicted the recognizability of melodies when presented a second time [20]. In these cases, it seems likely that working memory limitations make it difficult to encapsulate all aspects of a complex melody on first exposure. In summarizing recent research on working memory limitations, Cowan [1] proposes that there is a capacity of three to five chunks in working memory for young adults. According to these findings, modifications to the musical dimension bearing the least complex musical material should be the easiest to detect, which suggests that this musical dimension would have a predominant influence on similarity judgements. Collectively, this research supports the following hypothesis: modifications to the musical dimension bearing low complexity information will result in a significant decrease in similarity, in comparison to similar modifications to the musical dimension bearing high complexity information.

4. METHODOLOGY

4.1 Participants

The participants were recruited online using the Crowdflower [1] crowdsourcing platform, and required to pass a test before participating in the experiment. Participants were paid $0.02 USD for each question they answered, in accordance with the typical compensation offered to Crowdflower users. Of the 96 participants who took the test, 76 passed (79.2%) and 63 completed the experiment. 12 participants responses were deemed ineligible based on the inconsistent responses to an identical question. In total, 51 participants came from 25 different countries.

4.2 Stimuli

4.2.1 Measuring Complexity

Given the multifaceted nature of complexity, it is necessary to make the distinction between the entropy based complexity measures proposed by Eerola et al. [3], and the notion of complexity which grounds the current study. Shannon Entropy quantifies the disorder or uncertainty inherent in an information source based on a representative probability distribution [23]. Eerola et al. calculate entropy using the marginal probability of each symbol in a sequence. This type of complexity will be referred to as *entropy*$_m$. Although entropy$_m$ has been shown to correlate with the percieved complexity of musical sequences [16], this measurement of complexity does not provide the necessary resolution to make comparisons between many musical sequences. For an explicit example, consider the following pitch sequences, $s_1 = \{$c, d, e, f, c, d, e, f$\}$, and $s_2 = \{$c, f, e, d, e, c, d, f$\}$. Even though s_1 exhibits less complexity than s_2, both s_1 and s_2 have the same entropy$_m$, as this measurement does not take the repetition of longer phrases into consideration. Clearly, it is necessary to take the repetition of phrases into consideration when measuring complexity.

Admittedly, this can be accomplished by calculating the entropy rate of an n-th order markov chain derived from the musical sequence being measured, however there are still issues with this approach. In contrast to the manner in which humans percieve musical content, and by extension musical complexity, the entropy rate is not designed to distiguish between repetition which occurs within the prevailing metric structure, and repetition which spans metrical boundaries. Research suggests that humans perceive music by breaking it into a series of chunks [5], and have a natural tendency to project metre onto sequences of sound, despite the absence of acoustic cues for metric organization [4]. In addition, when listening to music, humans naturally extract motivic patterns [32], and larger formal structures [12]. Since humans segment music in accordance with metrical boundaries, it is likely that humans are less sensitive to repetition which is obscured by these boundaries. Consequently, a true measure of musical complexity must take this distinction into account.

Furthermore, an entropy based model of complexity is not capable of taking similarity into consideration, as entropy is based on the lossless encoding of an information source [23]. This becomes more of an issue when entropy is being measured with respect to larger subsequences, as is the case when measuring the n-th order entropy rate. This formulation of complexity cannot make the distinction between a collection of subsequences which share the same contour, and a collection that does not. As a result, it seems most reasonable to take the collective dissimilarity of subsequences segmented with respect to the prevailing metric structure, as a measure of complexity. Consequently, a homogenous collection of segments would be perceived as having a low complexity, while a diverse collection of segments would be perceived as having a high complexity. We use the term *redundancy* to refer to this type of complexity throughout the paper.

In order to quantify redundancy, two different measures were used. Thul's [28] adaptation of Tanguiane's [25, 26] algorithm, measures redundancy by counting the number of *root patterns*, at several hierarchical levels. This will be referred to as *Tanguiane's Rhythmic Complexity* (*TRC*). The other measure of redudancy is calculated using Eqn (1), where (S) is a set of subsequences, derived by segmenting a sequence of symbols into measures. Notably, Eqn (1) also requires a distance metric (D). Chronotonic distance [29] is used to measure *Rhythmic Sequence Complexity* (*RSC*), and a similarity measure proposed by Maidín [10] is used to measure *Pitch Sequence Complexity* (*PSC*). Admittedly, segmenting a pitch sequence according to metre means that *PSC* is dependant on the rhythmic content, however, within-measure rhythmic patterns have no bearing on *PSC* in this paradigm, and the metric structure is not being manipulated in this study. Although *PSC* does not account for the complexity of invidual segments, section 4.2.2 describes how complexity is restricted in this experiment, effectively mitigating the variance of segment complexity in the current study.

$$f(S) = \frac{1}{|S|}\sum_{i=1}^{|S|}\min\{D(S_i, S_j) : j \neq i; 1 \leq j \leq |S|\} \tag{1}$$

4.2.2 Prototype Melodies

In this experiment, there were four types of melodies; *rhythm*$_s$-*pitch*$_s$, *rhythm*$_s$-*pitch*$_c$, *rhythm*$_c$-*pitch*$_s$, and *rhythm*$_c$-*pitch*$_c$, where s denotes a simple or low complexity sequence, and c denotes a complex sequence [2]. In addition, eight versions of each melody type were constructed, resulting in 32 (4×8) prototype melodies of equal

[1] https://www.crowdflower.com/

[2] The melodies used in this experiment can be found at https://mlab%2Dexperiments.iat.sfu.ca/ismir2017/audio.

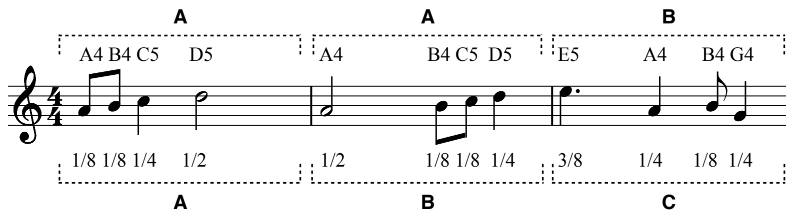

Figure 1. A melody with complex rhythm and simple pitch, using letters to show the form of each dimension.

length (three measures). As mentioned in section 4.2.1, redundancy quantifies the degree to which an information source is self similar and contains periodic repetition in conjunction with the prevailing metrical structure. In light of this aim, melodies were comprised of three measure-length phrases, with phrase repetition varied to create two distinct levels of complexity. Low complexity sequences had a formal pattern AAB, where a pattern is repeated in the first two measures, and a new pattern is introduced in the last measure. High complexity sequences had a formal pattern ABC, where each measure is dissimilar. This construction process is demonstrated in Figure 1, which shows a high complexity rhythm sequence and a low complexity pitch sequence.

Care was taken to restrict the variability of entropy$_m$ based complexity, using measures proposed by Eerola et al. [3]. Since the pitch sequences were constructed from scales consisting of five distinct pitch classes, *Entropy of pitch class distribution* and *Entropy of interval distribution* did not vary significantly. Similarly, rhythm sequences were constructed from four distinct durations, limiting the variance of *Entropy of note duration distribution* and *Rhythmic variability*. Notably, it seemed reasonable to have fewer distinct durations than pitch classes, as research has demonstrated that most listeners are able to perceive pitch diversity more readily [17]. A One-Way Analysis of Variance (ANOVA) across all four prototype melody types demonstrated that none of these entropy$_m$ based complexity measures were a significant source of variance, while *PSC*, *RSC* and *TRC* varied significantly. Furthermore, the entropy rate – calculated using a first order markov chain – did not vary significantly across melody type. This verified that our experiment measured the effect of variations in redundancy in relative isolation.

In order to restrict the variance of segment complexity, *Mean interval size* and *Note density* were restricted, which Eerola et al. [3] found to be a significant source of complexity. Each melody was constrained to an octave range, restricting the *Mean interval size*. The *Note density*, was invariant for each constructed melody, as each melody had four notes per measure, and was three measures long.

4.2.3 Modified Melodies

For each prototype melody (M), two modified versions were constructed for the main experiment: a version in which the pitch is modified $(\bar{M}_p)$, and a version in which the rhythm is modified $(\bar{M}_r)$. This process involved reversing the order of the measures in the dimension which is to be modified. As a result, regardless of the nature of the prototype melody, the first and last measures of the modified melody were different. Since test questions required a ground truth answer, three additional types of modified melodies were constructed: a melody in which the pattern form of M was transformed from AAB to ABA in the pitch dimension $(\bar{M}_{r\bar{p}})$, a melody in which the pattern form of M was transformed from AAB to ABA in the rhythm dimension $(\bar{M}_{p\bar{r}})$, and a melody in which both dimensions were modified $(\bar{M}_b)$.

4.3 Experimental Design

The experiment consisted of two independent variables, rhythm and pitch content complexity. Both rhythm and pitch complexity had two levels, low and high. This resulted in a 2×2 repeated measures experimental design, with four distinct types of prototype melodies. Participants were presented with a series of questions, consisting of a prototype melody (M) and two modified melodies (*melody A*, *melody B*). There were two types of test questions, which were developed using the modified melodies described above. The first type of question, compared either $\bar{M}_{r\bar{p}}$ and M against the prototype M, or $\bar{M}_{p\bar{r}}$ and M against M. This had an indisputable answer, as one of the modified melodies was in fact an exact replica of the prototype. The second type of question, compared $\bar{M}_p$ and $\bar{M}_b$ to the prototype, or compared $\bar{M}_r$ and $\bar{M}_b$ to the prototype. Given the manner in which these melodies were constructed, $\bar{M}_p$ and $\bar{M}_r$ are more similar to the prototype, as they are identical to the prototype along a single dimension, while $\bar{M}_b$ is dissimilar in both dimensions.

For the actual experiment itself, there was a single type of question, in which $\bar{M}_r$ and $\bar{M}_p$ were compared against the prototype. Irregardless of the type of question, the two modified melodies were randomly assigned to be *melody A* or *melody B*. For each question, participants rated the similarity of *melody A* to M, and *melody B* to M, on a Likert scale from 1 to 20, where 20 indicates maximal similarity. In the analysis below, the difference $(D = S(M, \bar{M}_r) - S(M, \bar{M}_p))$ between the perceived similarity of $\bar{M}_r$ to M $(S(M, \bar{M}_r))$, and the perceived similarity of $\bar{M}_p$ to M $(S(M, \bar{M}_p))$, is taken as the dependent variable. As a result, a positive value of D indicates that modifications to the rhythm dimension have less of an effect on similarity than modifications to the pitch dimension, while a negative value of D indicates the opposite.

4.4 Procedure

Before participating in the experiment, participants were required to complete 10 test questions with a minimum accuracy of 80%. The test questions served two purposes, eliminating those who were not taking the task seriously, and familiarizing participants with the similarity domain within which they were being asked to make comparisons. Once the test was successfully completed, participants were presented with 10 randomly ordered questions, consisting of eight different experiment questions (representing each of the eight different types of prototype melodies), a test question, and a repeated experiment question. The repeated experiment question was used to determine if participants were answering the questions consistently. For each question, the prototype melody was selected randomly from a collection of eight versions, and the key was randomly transposed so that the content varied from question to question. After listening to all three melodies, participants were asked to indicate which of the two modified versions was more similar to the prototype, and rate the similarity of *melody A* and *melody B* on a Likert scale from 1 to 20.

5. RESULTS

Since the ANOVA is relatively robust to violations of normality [21], the 2-Way ANOVA was conducted without transforming the data, despite the violation of the assumption of normality. A 2-Way ANOVA revealed the main effect of rhythm complexity ($F(50) = 9.17$, $p = .004$, $\eta_p^2 = .155$) and pitch complexity ($F(50) = 5.31$, $p = .025$, $\eta_p^2 = .096$), while the interaction between rhythm complexity and pitch complexity was insignificant ($p = .657$). To be thorough, an Aligned Rank Transform was performed on the data, correcting for the effects of the non-normal distributions of the data [33]. Using the transformed data, a 2-Way ANOVA revealed main effect of rhythm complexity ($F(50) = 9.82$, $p = .003$, $\eta_p^2 = .164$) and pitch complexity ($F(50) = 6.26$, $p = .016$, $\eta_p^2 = .111$), while the interaction between rhythm complexity and pitch complexity was insignificant ($p = .601$). These results corroborate the analysis of the untransformed data, indicating that 16.4% of the variability in similarity ratings were explained by changes in rhythm complexity, and 11.1% of the variability was explained by changes in pitch complexity.

As predicted, there was a main effect of rhythm complexity and pitch complexity, both shown in Figure 2b. Melodies containing low complexity rhythmic content ($M = 0.451$, $SD = 5.26$) were significantly lower than those containing high complexity rhythmic content ($M = 2.49$, $SD = 5.81$), which indicates that participants were more sensitive to pitch modifications when pitch sequences were less complex. This effect was pronounced in cases where the rhythmic sequence was more complex, as participants found pitch modified melodies ($\bar{M}_p$) to be significantly less similar to *rhythm*$_c$-*pitch*$_s$ prototype melodies than rhythm modified melodies ($\bar{M}_r$). Conversely, melodies containing low complexity pitch content ($M = 2.26$, $SD = 5.43$) were significantly higher than those containing high complexity pitch content ($M = 0.676$, $SD = 5.73$), which indicates that participants were more sensitive to rhythmic modifications when rhythmic sequences were less complex. Similarly, this effect was pronounced in cases where the pitch sequence was more complex, as participants found rhythm modified melodies ($\bar{M}_r$) to be significantly less similar to *rhythm*$_s$-*pitch*$_c$ prototype melodies than pitch modified melodies ($\bar{M}_p$). Therefore, the dimension bearing low complexity musical content was found to play a significant role in similarity judgements, as modifications to that dimension significantly decreased perceived similarity.

An analysis of the individual prototype melody conditions revealed that the *rhythm*$_s$-*pitch*$_c$ condition ($M = -0.235$, $SD = 5.21$) was significantly less than the *rhythm*$_c$-*pitch*$_s$ condition ($M = 3.39$, $SD = 5.39$), as pitch modified melodies were the most similar to *rhythm*$_s$-*pitch*$_c$ prototypes, and rhythm modified melodies were the most similar to *rhythm*$_c$-*pitch*$_s$ prototypes. The *rhythm*$_s$-*pitch*$_s$ condition ($M = 1.14$, $SD = 5.27$) and the *rhythm*$_c$-*pitch*$_c$ condition ($M = 1.59$, $SD = 6.12$) were roughly equivalent, and participants did not find a particular type of modified melody to be more similar, relative to the two other conditions. Collectively, these results indicate that melodies which are modified in the dimension bearing low complexity information are perceived as significantly less similar than melodies which are modified in the dimension bearing high complexity information.

6. DISCUSSION

As evidenced by the results presented above, modifications to the dimension bearing low complexity information result in a significant decrease in perceived similarity, demonstrating that the dimension bearing low complexity information plays a more significant role in melodic similarity judgments. On a whole, the values for all four conditions were positively skewed (Figure 2a), indicating that modifications to the pitch content of a melody had a greater influence on perceived similarity. Since there is no benchmark with which to compare rhythmic sequence complexity and pitch sequence complexity, it was not possible to equate the complexity across dimensions. Consequently, some skew in either direction was expected. The positive skew may indicate that the rhythmic content of the melodies in this experiment was on average more complex, and participants had difficulty noticing modifications in the rhythm dimension. Alternatively, due to the enculturation process that Hannon and Trehub [8] observed, participants may have paid more attention to the pitch content, resulting in the slight positive skew. When these factors are considered, it is arguably most meaningful to interpret the conditions in relation to each other, as some skew in either direction was inevitable. Viewed from this perspective, the hypothesis is directly corroborated, as the *rhythm*$_s$-*pitch*$_c$

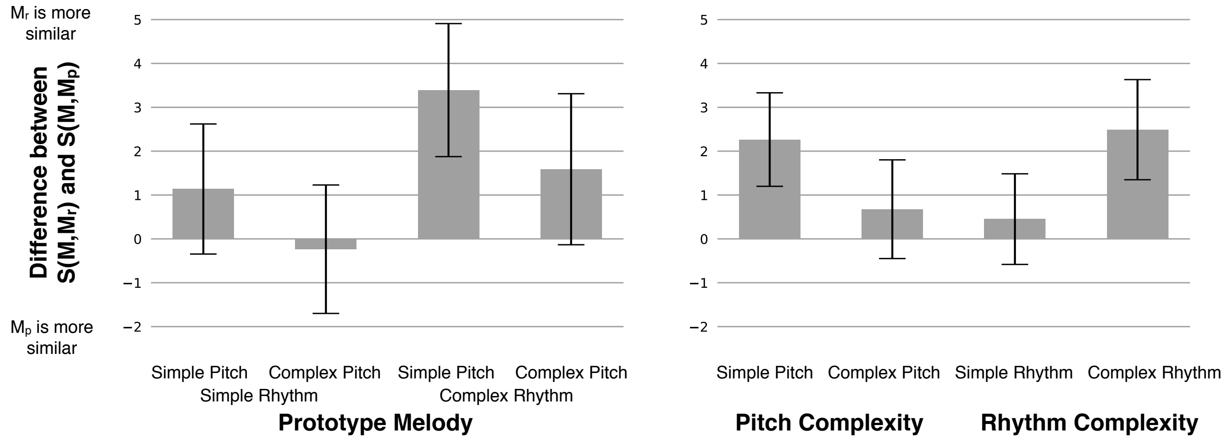

Figure 2. (a) The difference between the perceived similarity of the modified rhythm melody and the perceived similarity of the modified pitch version for each prototype melody complexity category, with 95% confidence intervals. (b) The main effects of pitch and rhythm complexity with 95% confidence intervals

condition is the lowest, the $rhythm_c$-$pitch_s$ is the highest, and the $rhythm_s$-$pitch_s$ and $rhythm_c$-$pitch_c$ conditions are in the middle.

Further analysis reveals that previous experiments are likely a special case of the generalized theory proposed in this paper. Monahan et al. [11] and Halpern [7] both make the claim that rhythm contributes more significantly to similarity perception, however, the rhythmic component of their stimuli is predominantly low complexity, and the pitch component of their stimuli is relatively higher on average. Notably, this was measured using *PSC*, *RSC*, and *TRC*. Although Halpern and Monahan et al. attribute their results to an inherent bias towards rhythm, the results of this experiment suggest that the relative complexity of the rhythm and pitch content provides a more robust explanation.

Admittedly, there are several limitations to the generalization of the results of this study. First and foremost, the observed relationship between dimensional complexity and similarity judgements may manifest itself quite differently when working with longer melodies, or polyphonic music. Secondly, due to the fact that musical complexity is multifaceted and far from understood, determining the relatively low complexity dimension may be quite difficult in some contexts. Despite the aforementioned limitations, the limited variance of Eerola et al.'s entropy based complexity measures provides substantial support for the generalization of these findings, as most western music makes use of the same limited collection of distinct note durations and pitch classes [16]. As a result, although this form of entropy based complexity is the source of some variability within the musical cannon, redundancy arguably accounts for more of this variation. Consequently, the results of this study are not restricted to a particular genre, and are relevant across musical genres.

7. CONCLUSION

Similarity is shaped by several factors, including familiarity, and cultural conditioning. This study asserts the significance of another factor – the nature of the musical content which is being compared – by examining the effects of dimensional complexity on similarity judgements. The general notion that characteristics of the musical content being compared have some bearing on the criterion used to make similarity judgements, is not new, and has been observed in past experiments [9]. However, the manner in which musical content establishes a criterion for similarity judgements has not been explored previously. The results of this study provide evidence that pitch and rhythmic complexity are factors which shape the criterion used in similarity judgements, as the dimension bearing relatively low complexity information has a greater influence on similarity perception. Furthermore, the results of this experiment are corroborated by previous experiments [7, 11], offering a general explanation for these previous findings.

Developing robust and flexible similarity measures continues to be a dominant area of research in the MIR domain, as large digital databases of music information necessitate accurate methods for comparison and categorization. As a result, adapting existing similarity measures to take dimensional complexity into account, is a possible application of the findings of this study. Future research is also necessary to investigate the role of complexity along other dimensions, including dynamics, articulation and timbre. Furthermore, the manner in which complexity is percieved along a single dimension is in need of continued exploration, as several issues with pre–existing methods for measuring complexity have been discussed in section 4.2.1. Clearly, musical similarity is a complex phenomenon which is deserving of continued exploration, as the results of this experiment have explicitly demonstrated that similarity judgements are dependant on another contextual factor, the complexity of pitch and rhythm content in the musical material being compared.

8. REFERENCES

[1] N. Cowan. The Magical Mystery Four: How Is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science*, 19(1):51–57, 2010.

[2] W. J. Dowling. Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85(4):341–354, 1978.

[3] T. Eerola, T. Himberg, P. Toiviainen, and J. Louhivuori. Perceived complexity of western and African folk melodies by western and African listeners. *Psychology of Music*, 34(3):337–371, 2006.

[4] P. Fraisse. Rhythm and Tempo. In Diana Deutsch, editor, *The Psychology of Music*, pages 149–181. Academic Press, New York, 1982.

[5] R. I. Godøy, A. R. Jensenius, and K. Nymoen. Chunking in music by coarticulation. *Acta Acustica united with Acustica*, 96(4):690–700, 2010.

[6] R. L. Goldstone. Learning to perceive while perceiving to learn. In R Kimchi, M Behrmann, and C Olson, editors, *Perceptual Organization in Vision: Behavioural and Neural Perspectives*, pages 233–278. Lawrence Erlbaum Associates, New Jersey, 2003.

[7] A. R. Halpern. Perception of structure in novel music. *Memory & cognition*, 12(2):163–170, 1984.

[8] E. E. Hannon and S. E. Trehub. Metrical categories in infancy and adulthood. *Psychological Science*, 16(1):48–55, 2005.

[9] A. Lamont and N. Dibben. Motivic Structure and the Perception of Similarity. *Music Perception: An Interdisciplinary Journal*, 18(3):245–274, 2001.

[10] D. O. Maidín. A geometrical algorithm for melodic difference. *Computing in musicology: a directory of research*, (11):65–72, 1998.

[11] C. B. Monahan and E. C. Carterette. Pitch and Duration as Determinants of Musical Space. *Music Perception: An Interdisciplinary Journal*, 3(1):1–32, 1985.

[12] C. Neuhaus, T. R. Knösche, and A. D. Friederici. Similarity and repetition: An ERP study on musical form perception. *Annals of the New York Academy of Sciences*, 1169:485–489, 2009.

[13] I. Peretz and M. Coltheart. Modularity of music processing. *Nature neuroscience*, 6(7):688–691, 2003.

[14] L. Pollard-Gott. Emergence of Thematic Concepts in Repeated Listening to Music. *Cognitive psychology*, 15(1):66–94, 1983.

[15] J. B. Prince. Contributions of pitch contour, tonality, rhythm, and meter to melodic similarity. *Journal of experimental psychology. Human perception and performance*, 40(6):2319–37, 2014.

[16] J. B. Prince and P. Q. Pfordresher. The role of pitch and temporal diversity in the perception and production of musical sequences. *Acta Psychologica*, 141(2):184–198, 2012.

[17] J. B. Prince, M. A. Schmuckler, and W. F. Thompson. The effect of task and pitch structure on pitch-time interactions in music. *Memory & cognition*, 37(3):368–381, 2009.

[18] T. W. Reiner. Pitch-distance and contour complexity in the recognition of short melodies. *Journal of Scientific Psychology*, (September):27–36, 2011.

[19] B. S. Rosner and L. B. Meyer. The perceptual roles of melodic process, contour, and form. *Music Perception*, 4(1):1–39, 1986.

[20] P. A. Russell. Memory for music: A study of musical and listener factors. *British Journal of Psychology*, 78(3):335–347, 1987.

[21] E. Schmider, M. Ziegler, E. Danay, L. Beyer, and M. Bühner. Is It Really Robust?: Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology*, 6(4):147–151, 2010.

[22] E. Schubert and C. Stevens. The effect of implied harmony, contour and musical expertise on judgments of similarity of familiar melodies. *Journal of New Music Research*, 35(2):161–174, 2006.

[23] C.E. Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27(July):379–423, 1948.

[24] C. Taher, R. Rusch, and S. McAdams. Effects of Repetition on Attention in Two-Part Counterpoint. *Music Perception*, 33(3):306–318, 2016.

[25] A. S. Tanguiane. *Artificial Perception and Music Recognition*. Springer-Verlag, Berlin, 1993.

[26] A. S. Tanguiane. A Principle of Correlativity of Perception and Its Application to Music Recognition. *Music Perception: An Interdisciplinary Journal*, 11(4):465–502, 1994.

[27] W. F. Thompson, M. D. Hall, and J. Pressing. Illusory conjunctions of pitch and duration in unfamiliar tone sequences. *J Exp Psychol Hum Percept Perform*, 27(1):128–140, 2001.

[28] E. Thul. *Measuring the Complexity of Musical Rhythm*. PhD thesis, McGill University, 2008.

[29] G. T. Toussaint. A comparison of rhythmic similarity measures. *Proc. International Conference on Music Information Retrieval (ISMIR 2004)*, pages 242–245, 2004.

[30] A. Volk, E. Chew, E. Hellmuth Margulis, and C. Anagnostopoulou. Music Similarity: Concepts, Cognition and Computation. *Journal of New Music Research*, 45(3):207–209, 2016.

[31] A. Volk and P. Kranenburg. Melodic similarity among folk songs: An annotation study on similarity-based categorization in music. *Musicae Scientiae*, 16(3):1–23, 2012.

[32] R. L. Welker. Abstraction of Themes from Melodic Variations. *Journal of Experimental Psychology. Human Perception and Performance*, 8(3):435–447, 1982.

[33] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 143–146, 2011.