

Perception of Rhythmic Similarity is Asymmetrical, and is Influenced by Musical Training, Expressive Performance, and Musical Context

Short title: Perception of Rhythmic Similarity

Cameron, Daniel; Potter, Keith; Wiggins, Geraint and Pearce, Marcus

Abstract

Rhythm is an essential part of the structure, behaviour, and aesthetics of music. However, the cognitive processing that underlies the perception of musical rhythm is not fully understood. In this study, we tested whether rhythm perception is influenced by three factors: musical training, the presence of expressive performance cues in human-performed music, and the broader musical context. We compared musicians and nonmusicians' similarity ratings for pairs of rhythms taken from Steve Reich's *Clapping Music*. The rhythms were heard both in isolation and in musical context and both with and without expressive performance cues. The results revealed that rhythm perception is influenced by the experimental conditions: rhythms heard in musical context were rated as less similar than those heard in isolation; musicians' ratings were unaffected by expressive performance, but nonmusicians rated expressively performed rhythms as less similar than those with exact timing; and expressively-performed rhythms were rated as less similar compared to rhythms with exact timing when heard in isolation but not when heard in musical context. The results also showed asymmetrical perception: the order in which two rhythms were heard influenced their perceived similarity. Analyses suggest that this asymmetry was driven by the internal coherence of rhythms, as measured by normalized Pairwise Variability Index (nPVI). As predicted, rhythms were perceived as less similar when the first rhythm in a pair had greater coherence (lower nPVI) than the second rhythm, compared to when the rhythms were heard in the opposite order.

Keywords: Rhythm perception, Similarity, Music cognition, Minimalism

1. Introduction

1.1 Rhythm

Rhythm is an essential part of musical experience. It allows us to move in synchrony with music and other listeners, it distinguishes musical styles and cultures, and it subtly guides our attention and expectations in time. Musical rhythm exists in all known human cultures, and the ability to move in time with musical rhythm occurs in humans without specialized training. However, there are wide individual differences in rhythm perception, associated with age (McAuley et al., 2006), culture (Soley & Hannon, 2010; Cameron et al., 2015), auditory short-term memory (Grahn & Schuit, 2012), and musical training (Palmer & Krumhansl, 1990; Kung et al., 2011; Bailey & Penhune, 2010; Chen et al., 2008). Trained musicians have more detailed representations of metrical structure (Palmer & Krumhansl, 1990) and more accurate perception of metrical structure and rhythmic groups than nonmusicians (Kung et al., 2011). Percussionists (a subset of musicians specializing in rhythm) have superior abilities for reproducing rhythms that are both beat-based and nonbeat-based, and for maintaining the beat with complex rhythmic sequences, compared to nonpercussionists (Cameron & Grahn, 2014). Thus, musical training seems to influence the neural processing of musical rhythm, although this has usually been demonstrated by superiority in task performance, and the influence of musical training on subjective measures of rhythm perception is less well understood.

Musical rhythm is associated with perceptual phenomena such as grouping (of auditory events), perception of regular and emphasized 'beat' (embedded within a dynamic, non-isochronous rhythm), and perception of hierarchical strong and weak beats (in a metrical structure). Musical rhythm is also associated with aesthetic appreciation, stylistic distinction, and the facilitation of precise timing and synchronization of motor actions (see London, 2004). Studies of the neural mechanisms and cognitive processing underlying musical rhythm perception indicate the important role that cortical and subcortical motor systems play in rhythm perception (e.g., Grahn & Brett, 2007; Chen et al., 2008; Kornysheva et al., 2010), and that both neural measures (e.g., induced beta-band activity; Fujioka et al., 2012) and cognitive measures (e.g., attention and expectation; Jones & Boltz, 1989) are modulated dynamically over time. Experimental studies of rhythm perception often use synthesized auditory sequences as stimuli, which lack expressive performance cues (i.e., performed variations in dynamics, timing and timbre), which are common in real music. Of particular interest here are expressive timing variations made by the performer. One study showed that tapping the beat with expressively timed music was less synchronized than with mechanically timed music, but that higher levels of the metrical hierarchy (slower rates) were tapped more often for expressively timed music, suggesting that expressive timing may contribute important information about the metrical structure of rhythm (Drake, et al., 2000). Repp (1998) demonstrated that when listening to music, listeners expect timing variations that are consistent with those heard in expressively timed music. Taken together, these studies suggest that expressive timing influences the cognitive processing of musical rhythm.

Experimental studies of rhythm perception also typically present individual rhythms in isolation or in repetition, rather than occurring within a broader musical context in which particular rhythms are intentionally chosen by a composer or performer to occur in a particular order. As a musical piece unfolds over time, the preceding rhythmic context is likely to influence perception of subsequent rhythms, but experimental research has yet to examine such an effect. The present study considers the influences of musical training, expressive performance, and musical context on perception of rhythms from a piece of music, *Clapping Music*, by Steve Reich.

1.2 Similarity

Perceptual similarity is a useful metric for investigating implicit processing of stimuli as it relies on individuals' intuitive categorization of perceptual phenomena whose nature and boundaries may not be explicitly accessible (Goldstone & Son, 2005). Using similarity perception to study musical rhythm perception is appropriate because rhythm-related behaviour (such as moving with the beat) and aesthetic appreciation (such as the sense of a rhythm's 'groove' or stylistic adherence) occur for virtually all humans including those without musical training or explicit knowledge of music theory and structure (see Phillips-Silver et al., 2011, regarding possible exceptions).

Various theoretical and computational approaches to perceptual similarity have been proposed. These include algorithmic, transformational approaches (e.g., Chater & Vitanyi, 2003), spatial/geometrical representations (e.g., Nosofsky, 1986; Shepard, 1987; Gardenfors, 2000), and feature-based, set-theoretic accounts (e.g., Tversky,

1977). In music cognition research, frequency-based statistics of features have been shown to account for some of the variability in ratings of perceptual similarity (Eerola et al., 2001). Another study compared expert listener's ratings of melodic similarity to the predictions of various computational models, in order to better understand the factors underlying cognitive representations of music (Müllensiefen and Frieler, 2004). Musicians' and nonmusicians' similarity ratings for piano music excerpts, while globally similar, were found to indicate underlying differences in processing: nonmusicians were more influenced by the characteristics of the music at the end of the excerpt compared to musicians who tended to use information from the entire excerpt (Lamont & Dibben, 2001).

Although computational approaches in music information retrieval have used similarity measures to study rhythm (e.g., Smith, 2010), these have not been applied to human perception. One study did use perceptual similarity ratings to validate a model of rhythmic similarity that uses a duration-based representation of rhythm (Hofmann-Engl, 2002). However, this research used only a limited number of simple sequences, and did not take these from real music. Forth (2012) demonstrated a metrical similarity model using a Gärdenfors (2000) conceptual space, whose distances corresponded with music-theoretic similarity between time signatures; again, this work did not empirically consider real music.

1.3 Asymmetrical Perception

Some theories of perception predict that the order in which two items are presented influences their similarity, and others do not. That is, depending on the computational approach, or factors considered, the perceptual similarity between two items may or may not be assumed to be the same when presented A-B vs. B-A. Set-theoretic and transformational models naturally predict asymmetry: both models assume that one perceived item can influence the perception of the subsequent item. Previous research using these approaches has explored asymmetry of human similarity judgments. Tversky (1977) discusses the issue from a theoretical perspective, and others extended his approach empirically (see Ortony et al., 1985), in work suggesting that feature-based contrasts of items can account for asymmetrical similarity. Standard geometric approaches preclude perceptual asymmetry (i.e., order-based differences in similarity). However, Nosofsky (1991) shows how many empirical cases of asymmetry in proximity data can be simulated in terms of a symmetric geometric model combined with bias components or weights associated with the salience of individual stimulus dimensions (see also Gärdenfors, 2000).

Asymmetry has been reported in perception of tonal stability such that pairs of chords or tones are perceived as more similar if the more stable one appears second (Bharucha & Krumhansl, 1983; Krumhansl, 1983). Furthermore, Bharucha and Pryor (1986) showed that recognition memory for two related auditory rhythmic sequences is better when the first is metrically coherent (defined using the framework of Povel, 1984, and Povel and Essens, 1985) and the second is a perturbed version of the first, compared to when presented in the opposite order. Dalla Bella and Peretz (2005) found that participants rated pairs of excerpts of Western classical music as more similar the closer

they were stylistically in terms of historical period. Interestingly, the styles were rated as less similar when presented in historical order, the older style preceding the more recent style (e.g. Baroque followed by Romantic), than the reverse (e.g. Romantic followed by Baroque). Rhythmic variability (measured by normalized pairwise-variability index, or nPVI; Patel & Daniele, 2003) was a strong predictor of the historical distance effect and Dalla Bella and Peretz (2005) suggest that the order effects may have been driven by earlier styles having lower rhythmic variability than later styles.

These results point towards a general cognitive effect whereby two stimuli are judged as being less similar if the first is more coherent (e.g., in terms of lower rhythmic variability), perhaps because it forms a stronger representation in memory and therefore serves as a better reference point for making a comparison than the less coherent stimulus (Bharucha & Pryor, 1986; Dalla Bella & Peretz, 2005). Therefore, it is important to investigate whether similar asymmetry exists in the perception of rhythmic similarity. We hypothesise that similarity ratings will be lower when the more coherent of a pair of rhythmic patterns (i.e., the one with lower nPVI) is presented first than when it is presented second.

1.4 Steve Reich's Clapping Music (1972)

“*Clapping Music*” (1972), by Steve Reich, is a standard piece of music from the minimalist repertoire (Potter, 2000). The piece requires two people to produce rhythms by clapping their hands, and has the following structure: both performers begin clapping

in unison a repeating rhythmic figure consisting of 12 isochronous units (a unit is a temporal position in which either one performer's clap, both performer's claps, or no claps can occur). The figure is repeated 12 times, at which point one of the performers shifts the rhythmic pattern ahead by one unit (time position), such that this performer starts on the second position, relative to the other performer, who continues with the original rhythm unchanged throughout the piece. In total, 12 rhythmic figures are performed and for each one the second performer shifts by one temporal unit relative to the previous figure. After the 12th figure, the final shift brings the two performers back into phase again, such that the 13th figure repeats the first figure. The present research considers only the first 12 distinct figures. Each figure is repeated 12 times. Thus, there are 12 unique rhythmic patterns that result from the discrete changes in phase relation between the two performers. See Fig.1 for a depiction of the rhythms in *Clapping Music*.

Ⓐ	1	2	3	4
Ⓑ	
RESULT	
Ⓐ	5	6	7	8
Ⓑ	
RESULT				.				.
Ⓐ	9	10	11	12
Ⓑ	
RESULT		.		.		.		

Figure 1. The twelve stimulus rhythms, taken from Steve Reich's *Clapping Music*. Lines indicate claps and dots indicate rests. For all twelve rhythms, A indicates the rhythm clapped by performer 1, and B indicates the rhythm clapped by performer 2. 'Result' indicates the overall rhythm resulting from the combination of the two performer's clapped rhythms. Intensity and timbral variation are not displayed, however, it is clear when both performers are to be clapping simultaneously, and when only one performer claps.

1. 5 Rhythmic Complexity in Clapping Music

Toussaint (2013) considers nPVI as a measure of rhythmic complexity, comparing it to an alternative measure, the standard deviation of the IOIs in a rhythm. Both measures are applied to a wide range of corpora ranging from rhythms artificially constructed for empirical studies (Essens, 1995; Fitch & Rosenfeld, 2007; Povel & Essens, 1985) to rhythms taken from African, Asian and European musical cultures. For some, but not all, corpora IOI standard deviation is significantly correlated with nPVI. nPVI also predicts rhythm reproduction performance in one case (Fitch & Rosenfeld, 2007) but not two others (Povel & Essens, 1985; Essens, 1995). nPVI differs significantly between the corpora representing distinct musical cultures and genres but does not distinguish between rhythms in a binary or ternary metre perhaps not surprisingly since it is a measure of rhythmic complexity rather than metrical complexity.

Toussaint (2013) points out that IOI standard deviation is completely blind to the order in which IOIs appear in a rhythm. While nPVI improves significantly on this measure, it is still oblivious to the underlying metre. Therefore, Toussaint proposes a *modified nPVI* which operates on the union of a metre and its underlying pulse (i.e., adding an audible metronome to the rhythm). It is not possible to compute modified nPVI since there is no time signature in the musical score and the piece is metrically ambiguous, potentially implying either a compound (12/8 or 6/2) or ternary (3/4) metre. Therefore we compare nPVI and IOI standard deviation in the analyses reported below.

1.6 Hypotheses

We propose four primary hypotheses: first, that musicians, due to their training, have enhanced sensitivity to subtle differences between individual rhythms, facilitating perception of rhythmic dissimilarity. Thus, we predicted that musicians would rate rhythms as less similar overall, compared to nonmusicians.

Second, we hypothesise that because human performance of musical rhythms includes expressive performance (subtle deviations from the exact rhythms as notated), it would also enable discriminability of rhythms. We thus predicted that expressively-performed rhythms would be rated as less similar to one another than rhythms with exact timing.

Third, we hypothesised that musical context contributes an extra dimension to the cognitive processing of rhythms, and thus predicted that rhythms heard in their musical context would be rated as less similar to one another, compared to rhythms heard in

isolation. We predict that the relationship formed between a rhythm and its preceding context constitutes an extra cognitive dimension, with its own unique properties, which are not available when the rhythm is heard in isolation.

Fourth, we hypothesised that similarity ratings for rhythms would be asymmetrical—that the order in which two rhythms were presented could influence their perceived similarity, and that this asymmetry would relate to the internal coherence of individual rhythms. As discussed above, we predict that when a rhythm with relatively greater coherence precedes a rhythm with relatively less coherence, the rhythms will be perceived as less similar, compared to when heard in the reverse order.

2. Methods

2.1 Participants

20 musicians (14 male, 6 female; mean age 26.5, SD = 7.02 years) and 20 nonmusicians (4 male, 16 female; mean age = 24.95, SD = 2.84 years) were recruited in London, UK. Group membership (musician or nonmusician) was confirmed by scores on the musical training subscale of the Goldsmiths Musical Sophistication Index (GMSI, (Müllensiefen et al., 2014): Musicians had a mean score of 50.95 (SD = 6.90) and nonmusicians a mean score of 25.70 (SD = 6.04), out of a maximum of 63.

Before testing, participants described their familiarity with minimalist music in order to exclude those who might have been familiar with *Clapping Music*. None were excluded on this basis. After testing, participants were asked if the stimuli sounded familiar, and if they were familiar with *Clapping Music*. None were.

2.2 Materials

Audio stimuli and visual instructions and cues were presented via laptop and Audio-Technica ATH-SJ3 stereo headphones.

Stimulus rhythms were drawn from a recording of Steve Reich's *Clapping Music* (1972). Two versions of the piece were used in this experiment. The performed version was an audio recording of two live performers performing the piece (Reich, 1980; Reich, 1972). The MIDI version was created programmatically by combining the rhythms performed by each performer, quantized so as to avoid any expressive variation in timing. The MIDI version was rendered to audio using six distinct individual 'clap' sounds, sampled from the performed recording for each performer (either performer 1, 2 or both) distinguishing whether or not the clap occurred at the first position in the metrical cycle (the downbeat), since claps in these positions are explicitly intended by the composer to be emphasized (Reich, 1980). Using these sounds controlled for differences in terms of timbre and intensity of the clap sounds between the performed and MIDI versions. The MIDI version is non-expressive in that timing, timbre and intensity of the clap do not vary between the 12 individual rhythms making up the piece.

2.3 Procedure

Participants completed the GMSI musical training subscale before beginning the similarity rating tasks. Participants were given a verbal description of the task and instructed to close their eyes while listening. Participants were instructed to rate rhythm-pairs on a scale from 1 to 7, with 1 being "minimum similarity" and 7 being "maximum similarity". It was emphasized that there were no correct or incorrect responses, or 'solutions', and that the intention was to collect intuitive judgments about similarity.

For the first task, participants rated the similarity of paired, isolated rhythms. First, the display read "Please close your eyes and listen" for 2 seconds, followed by a fixation cross at the centre of the screen at onset of the first rhythm, lasting 2.25s. The second rhythmic stimulus, of the same duration, was presented following 1.5s of silence. After the audio stimuli, the monitor displayed "On a scale from 1-7, with 7 being maximum similarity, how similar were the two rhythms you just heard?" A response triggered the next trial of a new pair of rhythms. For each participant, rhythm pairs were presented in randomized order and drawn from one of four stimulus subsets of all possible rhythm-pairs. Each

stimulus subset contained 78 rhythm pairs, each a combination of individual rhythms in one of the two possible versions (performed and MIDI) and one of the two possible orders (A-B and B-A). Half of the rhythm pairs in each subset were MIDI and half were performed. Each participant completed the task sitting at a desk in a quiet, isolated room, in approximately 12-15 minutes.

For the second task, participants heard progressively longer excerpts of *Clapping Music*. The first trial consisted of the first and second rhythmic figures, each repeated 4 times and presented consecutively; the second trial consisted of the first, second and third rhythmic figures, and so on. Each rhythmic figure was repeated 4 times instead of the 12, as originally intended by the composer, in order to keep testing sessions to a reasonable time. After each trial, the participant was asked to rate the similarity between the last rhythm in the trial and each of the previous rhythms, separately. Thus, for the first trial, one similarity rating was made (for the similarity of the first and second rhythmic figures); for the second, two ratings were made (between rhythmic figures 3 and 1, and between figures 3 and 2). If participants were unable to remember the earlier of the two, they could omit a response. Overall, 16.2% of ratings were omitted. Musicians and nonmusicians did not significantly differ in the proportion of missed ratings ($p = .68$). Each participant completed this procedure twice, once for each version (MIDI and performed). The order of versions was counterbalanced across participant groups, separately for musicians and nonmusicians.

Taken together, the two tasks yield a rating of similarity for each pair of rhythms by each participant in both individual paired presentation (Task 1) and within the musical context (Task 2).

2.4 Analyses

For the primary analysis, a 2x2x2 mixed design ANOVA was conducted on participants' mean similarity ratings for the repeated-measures factors of *Expressive Performance* (MIDI vs. performed versions of *Clapping Music*) and *Context* (isolated vs. in musical context), and the between-subjects factor of *Musical Training* (musicians vs. nonmusicians). Follow up *t* tests were conducted to test for differences between individual conditions in the case of significant interactions.

Further analyses were conducted to test whether any observed order effects might be related to rhythmic congruence between pairs of rhythms. For this purpose, rhythms are distinguished in terms of their nPVI score defined as follows:

$$nPVI = \frac{100}{m-1} \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1}) / 2} \right|,$$

in which m is the number of events in the rhythm and d_k is duration of the k th event (Patel & Daniele, 2003).

Specifically, we tested whether mean ratings (across both groups participants for both version conditions) correlate with the difference in nPVI scores between the individual rhythms in each pair. Pearson's correlation was used to examine relationships between similarity ratings and *absolute* nPVI differences (to demonstrate the validity of considering nPVI as a relevant factor in perception of rhythmic similarity). Pearson correlation was also used to examine *directional* nPVI differences (second rhythm minus the first) to test for a systematic relationship between perceived similarity and the order of rhythms' associated nPVI values.

In addition, we compared the mean ratings for each rhythm pair with a non-zero nPVI difference (averaged across participants) in a 2x2x2 repeated measures ANOVA with the factors *Musical Training* (musicians vs. nonmusicians), *Expressive Performance* (MIDI vs. performed), and *nPVI order* (whether the first or second rhythm had the greater nPVI).

The assignments of these variables as repeated measures factors (despite the fact that similarity ratings were averaged across participants) was justified because the exact same pairs of rhythms were being compared directly in the 8 (2x2x2) conditions. Again, because order effects could only apply to rhythms presented as pairs, this analysis was applied only to data from Task 1.

3. Results

3.1 Primary Analyses

Overall, the results contradict one of the primary hypotheses, corroborate three, and reveal interactions between the factors.

Contrary to the first hypothesis, there was no main effect of musical training: The average similarity ratings of musicians and nonmusicians did not significantly differ ($F(1,38) = 0.89, p = .350$). Consistent with the second hypothesis, there was a significant main effect of expressive performance; MIDI rhythm-pairs were rated as more similar compared to performed rhythm-pairs ($F(1,38) = 3.61, p = .033$, 1-tailed). However, these two factors interacted ($F(1,38) = 4.34, p = .044$), such that nonmusicians rated MIDI rhythm-pairs as more similar than performed rhythm-pairs ($t(19) = 3.01, p = .007$), but musicians did not rate the two types differently ($t(19) = 0.12, p = .904$).

As predicted by the third hypothesis, participants rated rhythm-pairs as less similar when they were heard in the context of the musical composition, than when heard in isolation (main effect of musical context, $F(1,38) = 17.81, p < .001$). There was also a significant interaction between musical context and expressive performance ($F(1,38) = 12.51, p = .001$), such that, when heard as isolated pairs, rhythms with expressive performance were rated as being less similar than rhythms without expressive performance ($t(39) = 3.51, p = .001$), but this effect of expressive performance was not present when rhythms were heard in musical context ($t(39) = 0.81, p = .422$).

For both musicians and nonmusicians, and both MIDI and performed versions of the rhythms, rhythm-pairs heard in musical context had lower mean similarity ratings than rhythm-pairs heard in isolation, as shown in Fig. 2 (musicians-MIDI, $t(19) = 4.13, p$

< .001; musicians-performed, $t(19) = 2.89, p = .009$; nonmusicians-MIDI, $t(19) = 6.16, p < .001$; nonmusicians-performed, $t(19) = 2.44, p = .025$).

To ensure that effects of Context were not mediated by poor memory for figures appearing further in the past, an additional analysis was conducted on the 16.2% of ratings that were omitted by participants. The proportion of omitted ratings did not correlate with the temporal distance between the two rhythms being compared ($r = .02, p = .85$). This suggests that memory demands associated with completing the task did not influence performance.

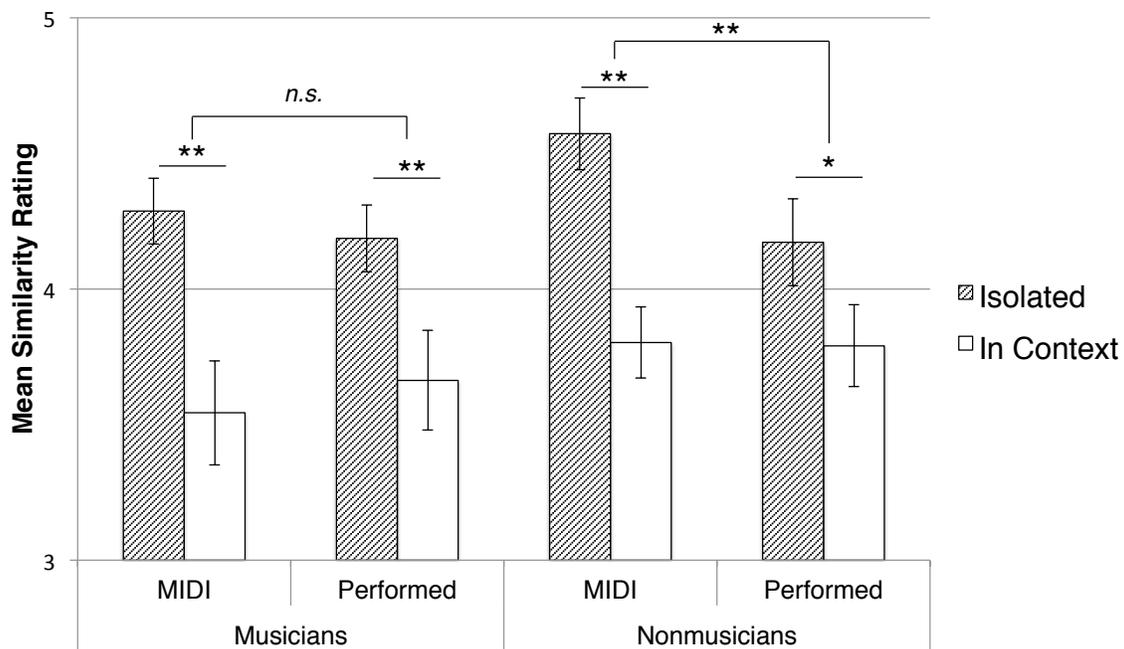


Figure 2. Mean similarity ratings of musicians and nonmusicians for rhythm pairs presented in isolation and in their musical context. Error bars indicate +/- 1 SEM. * indicates $p < .05$. ** indicates $p < .01$.

3.2 nPVI and Perceptual Asymmetry

Results from analyses considering nPVI and perceptual asymmetry support the fourth hypothesis. As shown in Fig. 3, absolute differences in nPVI between paired rhythms negatively correlated with similarity ratings ($r = -.686, p < .001$), indicating that nPVI captures an aspect of rhythm structure that contributes to the perception of rhythmic similarity (rhythms with more similar nPVI are perceived as more similar). Moreover, the directional differences in nPVI scores (nPVI of the second rhythm minus nPVI of the first rhythm) negatively correlated with similarity ratings ($r = -.179, p = .040$), indicating that the extent to which the second rhythm has lower nPVI (greater coherence) than the first is associated with the rhythms' perceived dissimilarity.

Results of the 2x2x2 repeated measures ANOVA including the factor *nPVI Order* showed that rhythm pairs in which the second rhythm had higher nPVI score than the first were rated as less similar than when the same rhythms were presented in the opposite order (main effect of nPVI order, $F(1,57) = 34.79, p < .001$). Additionally, this nPVI-based asymmetry depended on the musical training of the listener and the version (MIDI vs. Performed) of the rhythms (3-way interaction of *Musical Training*, *Expressive Performance*, and *nPVI Order*, $F(1,57) = 4.57, p = .037$). Follow-up paired *t* tests showed that the nPVI-based asymmetry influenced musicians' ratings of both MIDI and performed rhythms (MIDI: $t(57) = 3.69, p = .001$; Performed: $t(57) = 2.76, p = .008$), but for nonmusicians, nPVI-based asymmetry only had an influence for performed rhythms ($t(57) = 5.07, p < .001$) and not for MIDI rhythms ($t(57) = 1.30, p = .198$), as shown in Fig. 4.

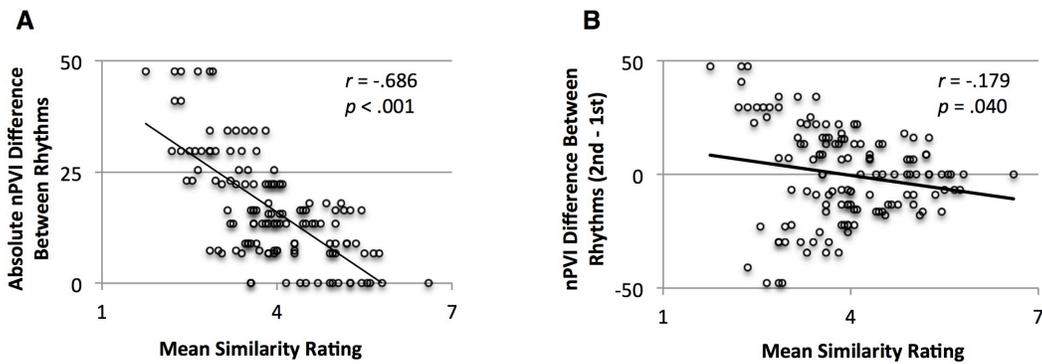


Figure 3. Correlations between mean similarity ratings for rhythm pairs and: A) absolute nPVI differences between rhythm pairs; B) directional nPVI differences (2nd rhythm minus 1st rhythm).

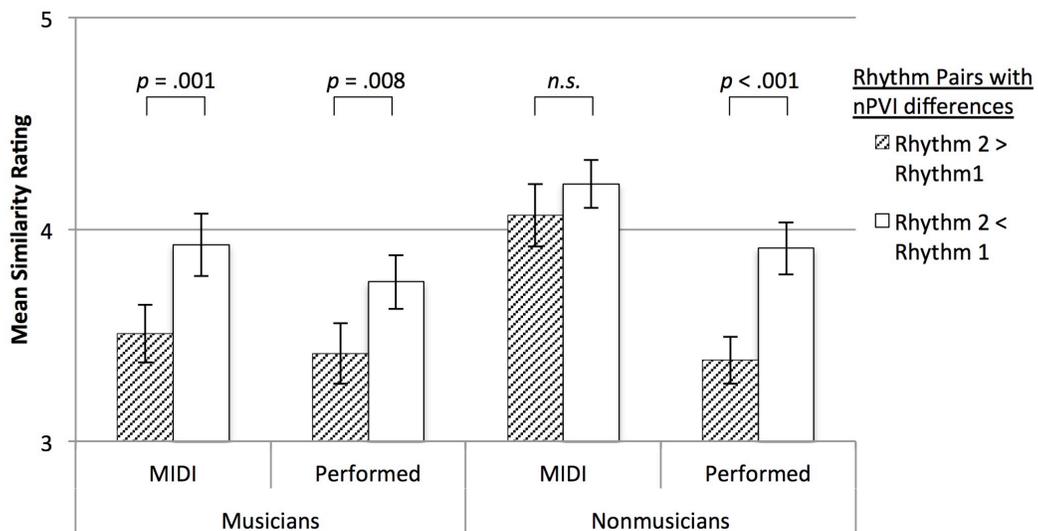


Figure 4. Mean similarity ratings of musicians and nonmusicians for trials in which the first rhythm had higher nPVI than the second. Error bars indicate +/- 1 standard error of the mean.

The analysis was then repeated using IOI standard deviation in place of nPVI (Toussaint, 2013). Across the 12 rhythmic figures in Clapping Music, there is a high correlation between nPVI and IOI standard deviation, $r(10) = .90, p < .001$. The analysis produced identical patterns of significant results with the exception that in the follow-up paired t tests, the effect of asymmetry on musicians' ratings of performed rhythms is marginally non-significant, $t(50) = 1.79, p = .08$.

4. Discussion

Overall, the results show that perception of musical rhythms (as measured by ratings of their similarity) is influenced by whether or not the listener is musically trained, by whether the rhythms include or lack expressive human performance cues, and by whether or not the rhythms are heard within a musical context. These factors also interacted with one another in their influence on perceived similarity of rhythms.

Nonmusicians rated MIDI rhythm-pairs (with mechanical timing) as being more similar than expressively performed rhythm-pairs. There are two possible interpretations of this finding. First, it may be that nonmusicians (but not musicians) gain information from the subtle timing variations of expressive human performance in judging similarity between individual rhythms. Second, it may be that nonmusicians exhibited a response bias such that MIDI rhythms tended to be rated as similar while performed rhythms tended to be rated as dissimilar. The two interpretations are not mutually exclusive. The fact that the effect appears to be driven primarily by nonmusicians' higher similarity ratings for MIDI stimuli compared to musicians (while mean ratings for performed

rhythms are comparable between the two groups) suggests that nonmusicians found the MIDI rhythms to be similar, requiring performance features to perceive stimuli as dissimilar to the extent that musicians do. For musicians, however, individual rhythms are judged to be dissimilar based on their temporal structure alone to a degree that expressive performance does not contribute. This result partially supports the first and second hypothesis, concerning the respective influences of musical training and expressive performance on rhythm perception. However, these influences appear to be more complex and inter-related than we had hypothesised suggesting that future research should focus on hypothesis-driven studies to replicate these interactions and corroborate the following explanations we provide.

Musicians' superior abilities to accurately organize rhythms into a hierarchical metrical organization may underlie the finding that, unlike nonmusicians, similarity ratings did not differ between human-performed and MIDI renditions (the latter lacking the subtle timing variation of expressive human performance). This is consistent with findings from previous studies of musicians and nonmusicians. Musicians are better able to perceptually organize music into a metrical hierarchy (Palmer & Krumhansl, 1990) and to use that hierarchy while synchronizing with music whether or not it contains expressive timing (Drake & Palmer, 2000). Musicians are also more sensitive than nonmusicians to both the presence and the absence of a beat in rhythms (Grahn & Rowe, 2009). For nonmusicians, we hypothesise that the effects of expressive performance are related particularly to expressive timing given the rhythmic nature of the stimulus but it is possible that expressive changes in loudness and timbre also have

an impact. Future research should examine this question using specially created stimuli that orthogonalise these different dimensions of expressive performance.

There is a co-linearity between musical training and sex in our sample, which means that it is possible that the effects of musical training could, in fact, reflect differences between men and women. However, as far as we can find in the literature there is no theoretical or empirical rationale for predicting sex differences in rhythm similarity perception whereas such a rationale does exist for musical training. Nonetheless, future research should discount this possibility explicitly.

Supporting the third hypothesis, participants rated rhythm pairs as less similar when they were heard in the context of the original piece of music than when heard in isolation. This underscores the importance of the larger structure in which musical rhythms are normally perceived. It seems likely that the relationship between a rhythm and its context provides an extra dimension with scope for additional unique properties, which is lacking when the rhythm is heard in isolation. This result suggests that care must be taken when generalizing the interpretation of results for the perception of stimuli isolated from their natural context and underlines the importance of complementing research using carefully controlled, artificial stimuli with studies using more naturalistic, ecologically valid stimuli.

The results suggest that musical rhythms are subject to asymmetrical perception, or order effects. Perceived similarity of rhythm pairs differed, depending on the order in which they were heard. We hypothesised that this effect may be driven by differences in coherence (measured by nPVI and IOI standard deviation in the present work) between the rhythms making up a pair. Specifically, we predicted that when the more coherent

rhythm was heard first, perceived similarity would be lower, due to more accurate encoding of the first, more coherent rhythm, facilitating its use as a reference for comparison with the second, less coherent, rhythm (Bharucha & Pryor, 1986; Dalla Bella & Peretz, 2005). The results support this hypothesis: the lower the coherence of the first rhythm presented relative to the second, the lower the perceived similarity between the rhythm pairs. Since nPVI and IOI standard deviation are highly correlated across the 12 rhythmic figures of Clapping Music, we were unable to distinguish them experimentally and the analysis produced almost identical results using either measure. This suggests that the results do not depend critically on the choice of nPVI as a measure of rhythmic coherence.

Moreover, when similarity ratings for rhythm pairs with non-zero nPVI differences were compared between each order of presentation, a main effect of nPVI-order showed that similarity was lower when the first rhythm has lower nPVI (i.e., greater coherence) than the second. However, the 3-way interaction of *Musical Training*, *Expressive Performance*, and *nPVI order* reveals that this asymmetry was not present for nonmusicians' perception of MIDI versions of rhythms. This may be due to the fact that nonmusicians, but not the musicians, showed higher mean similarity ratings for the MIDI versions than the performed versions of rhythms, as described above. That is, the perceptual benefit of hearing performed rhythms containing expressive performance, seems to facilitate perceptual asymmetry, and when additional discriminability of expressive performance is removed in the MIDI rhythms, asymmetry in nonmusicians' similarity ratings is eliminated. Future research should focus on

testing this explanation of the interaction between musical training and expressive performance.

It is also interesting to note that rhythm pairs were rated as more similar when they were presented in the order in which they occur in their original source, *Clapping Music* ($t(263) = 2.725, p < .01$). This is notable since the compositional process of the piece involves rotating (or phasing) one pattern with respect to another. In principle, the rotation could be achieved in two directions, with the effect that the 12 figures appear in reverse order. It is interesting that Reich chose to apply the phasing such that the order of patterns increases perceptual similarity between consecutive pairs of patterns, although whether this was a factor (implicit or explicit) in the compositional process is not documented. However, other research has also suggested relationships between order-related asymmetries in rhythm perception and compositional form. In a study of the perception of rhythmic stimuli that changed from unsyncopated to syncopated and syncopated to unsyncopated, Keller and Schubert (2011) found that only the former elicited perceived changes in complexity, relating this result to formal structure, such as theme and variation.

Overall, the analysis of asymmetry in the similarity ratings corroborates previous findings that pairs of stimuli are judged to be more similar when the first is more coherent than the second (Bharucha & Krumhansl, 1983; Bharucha & Pryor, 1986; Dalla Bella & Peretz, 2005; Krumhansl, 1983) and extends these results to purely rhythmic, real-world musical stimuli.

Taken together, the results of this study demonstrate that the perception and cognitive representation of musical rhythms, as indexed by similarity ratings, differs between musicians and nonmusicians, is influenced by expressive performance, and by the presentation of rhythm within a broader musical context. Furthermore, rhythm perception is asymmetrical, in that listeners perceive two rhythms as being less similar if the more coherent rhythm of the pair is presented first.

5. Acknowledgements

The authors would like to thank Jocelyn Bentley for her help with Fig. 1. Funding for this study was provided by EPSRC grant EP/H01294X/1.

6. References

- Bailey, J. A., & Penhune, V. B. (2010). Rhythm synchronization performance and auditory working memory in early- and late-trained musicians. *Exp Brain Res*, *204*(1), 91-101. doi:10.1007/s00221-010-2299-y
- Bharucha, J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: hierarchies of stability as a function of context. *Cognition*, *13*(1), 63-102.
- Bharucha, J. J., & Pryor, J. H. (1986). Disrupting the isochrony underlying rhythm: an asymmetry in discrimination. *Percept Psychophys*, *40*(3), 137-141.
- Cameron, D. J., Bentley, J., & Grahn, J. A. (2015). Cross-cultural influences on rhythm processing: reproduction, discrimination, and beat tapping. *Front Psychol*, *6*, 366. doi:10.3389/fpsyg.2015.00366

- Cameron, D. J., & Grahn, J. A. (2014). Enhanced timing abilities in percussionists generalize to rhythms without a musical beat. *Front Hum Neurosci*, 8, 1003. doi:10.3389/fnhum.2014.01003
- Chater, N., & Vitanyi, P.M.B. (2003). The generalized universal law of generalization. *Journal of Mathematical Psychology*, 47(3): 346-369.
- Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *J Cogn Neurosci*, 20(2), 226-239. doi:10.1162/jocn.2008.20018
- Dalla Bella, S., & Peretz, I. (2005). Differentiation of classical music requires little learning but rhythm. *Cognition*, 96(2), B65-78. doi:S0010-0277(05)00019-3 [pii] 10.1016/j.cognition.2004.12.005
- Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: Relations between planning and temporal control. *Cognition*, 74(1), 1-32.
- Drake, C., Penel, A., & Bigand, E. (2000). Tapping in time with mechanically and expressively performed music. *Music Perception: An Interdisciplinary Journal*, 18(1), 1-23.
- Eerola, T., Järvinen, T., Louhivuori, J., & Toiviainen, P. (2001). Statistical features and perceived similarity of folk melodies. *Music Perception*, 18: 275-296.
- Essens, P. (1995). Structuring temporal sequences: Comparison of models and factors of complexity. *Perception & Psychophysics*, 57(4), 519-532.
- Fitch, W. T., & Rosenfeld, A. J. (2007). Perception and production of syncopated rhythms. *Music Perception*, 25, 43-58.

- Forth, J. (2012). *Cognitively-motivated geometric methods of pattern discovery and models of similarity in music*. PhD thesis. Goldsmiths, University of London, UK.
- Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic β oscillations. *J Neurosci*, 32(5), 1791-1802. doi:10.1523/JNEUROSCI.4107-11.2012
- Gärdenfors, P. (2000). *Conceptual spaces: the geometry of thought*. A Bradford book. *MIT Press*, 3, 16.
- Goldstone, R. L., & Son, J. Y. (2005). *Similarity*. Cambridge University Press.
- Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *J Cogn Neurosci*, 19(5), 893-906. doi:10.1162/jocn.2007.19.5.893
- Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J Neurosci*, 29(23), 7540-7548. doi:10.1523/JNEUROSCI.2018-08.2009
- Grahn, J. A., & Schuit, D. (2012). Individual differences in rhythmic ability: Behavioral and neuroimaging investigations. *Psychomusicology: Music, Mind, and Brain*, 22(2), 105.
- Hofmann-Engl, L. (2002). Rhythmic similarity: A theoretical and empirical approach. In *Proceedings of the Seventh International Conference on Music Perception and Cognition* (pp. 564-567).
- Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychol Rev*, 96(3), 459-491.
- Keller, P. E., & Schubert, E. (2011). Cognitive and affective judgements of syncopated musical themes. *Advances in Cognitive Psychology*, 7, 142-156.

- Kornysheva, K., von Cramon, D. Y., Jacobsen, T., & Schubotz, R. I. (2010). Tuning-in to the beat: Aesthetic appreciation of musical rhythms correlates with a premotor activity boost. *Hum Brain Mapp*, *31*(1), 48-64. doi:10.1002/hbm.20844
- Krumhansl, C.L. (1983). Perceptual structures for tonal music. *Music Perception*, *1*, 24-58.
- Kung, S. J., Tzeng, O. J., Hung, D. L., & Wu, D. H. (2011). Dynamic allocation of attention to metrical and grouping accents in rhythmic sequences. *Exp Brain Res*, *210*(2), 269-282. doi:10.1007/s00221-011-2630-2
- Lamont, A., & Dibben, N. (2001). Motivic structure and the perception of similarity. *Music Perception*, *18*(3): 245-274.
- London, J. (2004). *Hearing in Time: Psychological Aspects of Musical Meter*. Oxford: Oxford University Press.
- McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: life span development of timing and event tracking. *J Exp Psychol Gen*, *135*(3), 348-367. doi:10.1037/0096-3445.135.3.348
- Müllensiefen, D., & Frieler, K. (2004). Cognitive adequacy in the measurement of melodic similarity: algorithmic vs. human judgements. *Computing in Musicology*, *13*: 147-176.
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PLoS One*, *9*(2), e89642. doi:10.1371/journal.pone.0089642
- Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, *23*(1), 94-140.

- Ortony, A., Vondruska, R. J., Foss, M. A., & Jones, L. E. (1985). Salience, similes, and the asymmetry of similarity. *Journal of memory and language*, 24(5), 569-594.
- Palmer, C., & Krumhansl, C. (1990). Mental representations for musical meter. *J Exp Psychol Hum Percept Perform*, 16(4), 728-741.
- Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, 87(1), B35-45.
- Phillips-Silver, J., Toiviainen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., & Peretz, I. (2011). Born to dance but beat deaf: a new form of congenital amusia. *Neuropsychologia*, 49(5), 961-969. doi:10.1016/j.neuropsychologia.2011.02.002
- Potter, K. (2000). *Four Musical Minimalists : La Monte Young, Terry Riley, Steve Reich, Philip Glass*. Cambridge, UK ; New York: Cambridge University Press.
- Povel, D. J. (1984). A theoretical framework for rhythm perception. *Psychol Res*, 45(4), 315-337.
- Povel, D., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, 2(4): 411-440.
- Reich, S. (1972). *Clapping Music*. Music recording. Nonesuch Records
- Reich, S. (1974). *Writings About Music*. New York: NYU Press.
- Reich, S. (1980). *Clapping Music*. Musical score. London: Universal Edition.
- Repp, B. H. (1998). Obligatory "expectations" of expressive timing induced by perception of musical structure. *Psychol Res*, 61(1), 33-43.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237(4820), 1317-1323.
- Smith, L. (2010). *Rhythmic similarity using metrical profile matching*. Ann Arbor, MI:

Michigan Publishing, University of Michigan Library.

Soley, G., & Hannon, E. E. (2010). Infants prefer the musical meter of their own culture: a cross-cultural comparison. *Dev Psychol*, *46*(1), 286-292. doi:10.1037/a0017555

Toussaint (2013). The Pairwise Variability Index as a Measure of Rhythm Complexity. *Analytical Approaches to World Music*, *2*(2).

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*: 327-352.