Title: Invariance of Edit-Distance to Tempo in Rhythm Similarity

Authors: Matthew Moritz[1], Matthew Heard[2], Hyun-Woong Kim[2], and Yune S. Lee[2]

Affiliations: [1]Department of Neuroscience, The Ohio State University, Columbus, OH 43210, USA. [2]School of Behavioral and Brain Sciences, The University of Texas at Dallas, Dallas, TX 75080, USA.

Corresponding Author: Yune S. Lee, Ph.D.

Address: Callier Center Richardson, 2895 Facilities Way, Richardson, TX 75080

Telephone: 603-443-0505

E-mail: yune.lee@utdallas.edu

Word Count: 4482

1. Abstract

Despite the long history of music psychology, rhythm similarity perception remains largely unexplored. Several studies suggest that edit-distance—the minimum number of notational changes required to transform one rhythm into another—predicts similarity judgements. However, the ecological validity of edit-distance remains elusive. We investigated if the edit-distance model can predict perceptual similarity between rhythms that also differed in a fundamental characteristic of music—tempo. Eighteen participants rated the similarity between a series of rhythms presented in a pair-wise fashion. The edit-distance of these rhythms varied from 1 to 4, and tempo was set at either 90 or 150 beats per minute. A test of congruence among distance matrices (CADM) indicated significant inter-participant reliability of ratings, and non-metric multidimensional scaling (nMDS) visualized that the ratings were clustered based upon both tempo and whether rhythms shared an identical onset pattern, a novel effect we termed rhythm primacy. Lastly, Mantel tests revealed significant correlations of edit-distance with similarity ratings on both within-tempo and between-tempo rhythms. Our findings corroborated that the edit-distance predicts rhythm similarity and demonstrated that the edit-distance accounts for similarity of rhythms that are markedly different in tempo. This suggests that rhythmic gestalt is invariant to differences in tempo.

Keywords: rhythm, similarity, edit-distance, tempo, primacy

2. Introduction

Rhythm, the temporal patterns of sound onsets, is an integral part of music structure and can provide a potent cue to song identification even without melodic or harmonic information. For example, an enthusiast of classical music could identify some of the most distinct compositions in classical music, such as Tchaikovsky's *1812 Overture*, Beethoven's *Fifth Symphony*, or *Mars, the Bringer of War* from Holst's *The Planets*, solely based upon rhythm. Outside the world of classical, jazz musicians often improvise main rhythmic themes, (re)forming an important part of both a song's and a musician's characteristics. Furthermore, composers can use rhythms that are similar to each other in order to tie in motifs, providing a sense of identity or togetherness for a piece of music. For computational purposes, rhythm similarity is also a crucial dimension for music database algorithms that classify songs within the same genre or category (Panteli, Bogaards, & Honingh, 2014; Paulus & Klapuri, 2002). As such, the psychological mechanisms and computational principles that underlie rhythm similarity have been queried by scholars in music theory, musicology, and psychology (Cao, Lotstein, & Johnson-Laird, 2014; Orpen & Huron, 1992; Post & Toussaint, 2011).

An early model of rhythm similarity (Toussaint, Matthews, Campbell, & Brown, 2012; Tversky, 1977) assessed similarity between rhythm phrases on the basis of shared features (Figure 1). Inspired by geometry, this feature-based model visually represented rhythms as circular, 2-dimensional shapes consisting of notes and rests as represented by black and white circles respectively (Figure 1). By connecting black dots in the circle, one can readily appreciate the rhythmic structure and extract distinct features (e.g., mirror symmetry). This, in turn, would help to discern the degree of similarity between different rhythm phrases. For example, two rhythms that are symmetrical in this diagram are expected to sound highly similar (e.g., R1 vs. R2 in Figure 1) compared to a rhythm without this feature (e.g., R1 vs. R3 in Figure 1).

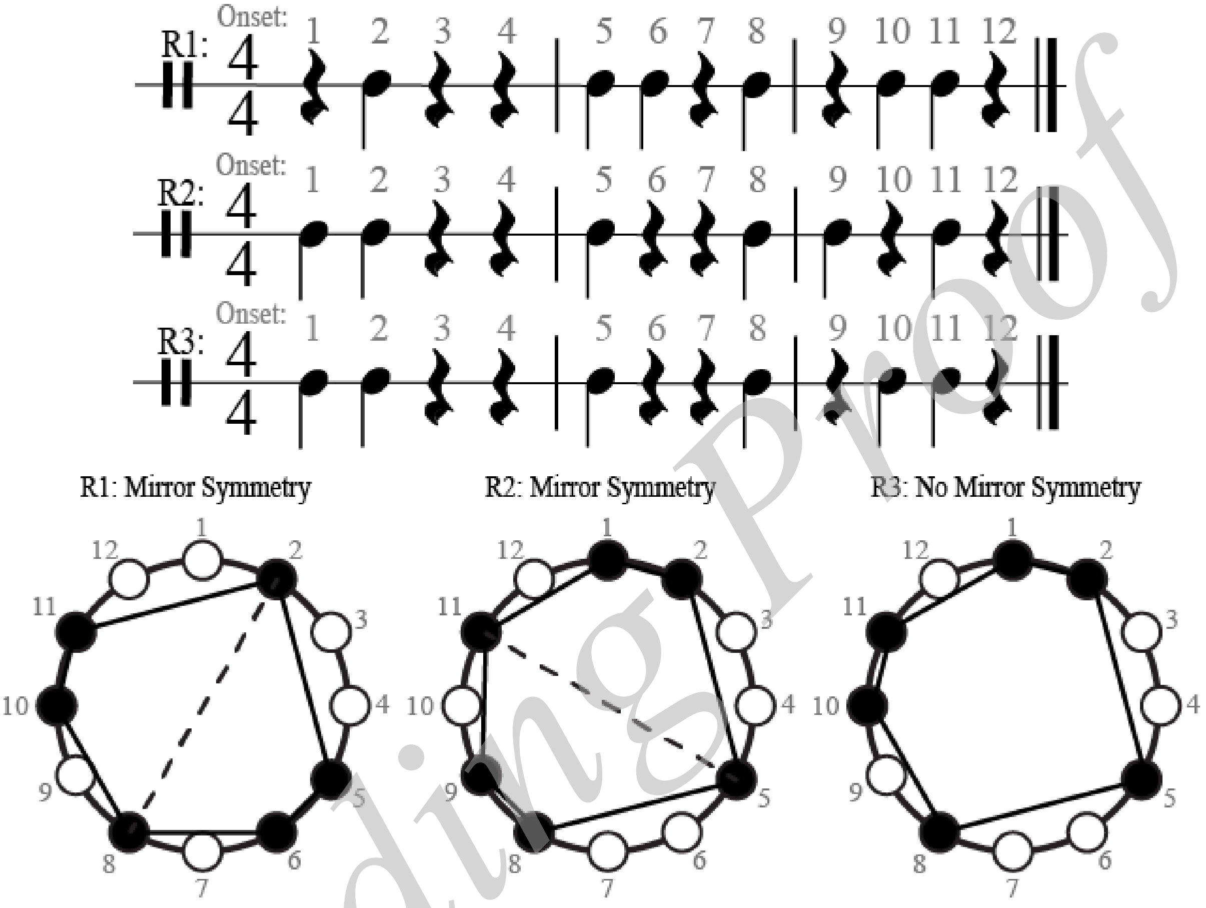

Figure 1. Three rhythm phrases written in both musical notation and geometric notation for feature extraction. Both rhythm 1 (R1) and rhythm 2 (R2) exhibit mirror symmetry about one axis, while rhythm 3 (R3) does not. Thus, feature-based theory postulates that R1 is more similar to R2 than R3 due to shared mirror symmetry. Adapted from Toussaint et al. (2012).

More recently, the edit-distance model eschewed this feature-based rhythm similarity account in favor of a transformational approach (Toussaint et al., 2012). Transformational approaches of similarity like edit-distance are used in many domains, for example to assess similarity between strings of character symbols in computer science (Lowrance & Wagner, 1975; Wagner & Fischer, 1974) as well as between melodic sequences using musical database search algorithms and string matching techniques (Cambouropoulos, Crawford, & Iliopoulos, 2001;

Typke, Veltkamp, & Wiering, 2004). Edit-distance is defined as the minimum number of edits—operationalized as insertions, deletions, and substitutions—of rhythm units required to transform one rhythm phrase into another (Figure 2). Fewer edits corresponds with a higher degree of rhythm similarity (Orpen & Huron, 1992; Post & Toussaint, 2011). Importantly, edit-distance was shown to be more successful at predicting human perception of rhythm similarity than feature-based approaches (Toussaint & Oh, 2016; Toussaint et al., 2012). Nevertheless, computational models of rhythm similarity often ignore ecological validity, and edit-distance is no exception. Prior studies of edit-distance are limited by their use of overly simple rhythmic patterns with identical tempos (Toussaint & Oh, 2016; Toussaint et al., 2012), naturally inviting an important question of whether or not edit-distance still accounts for perceptual similarity between rhythms of different tempos.

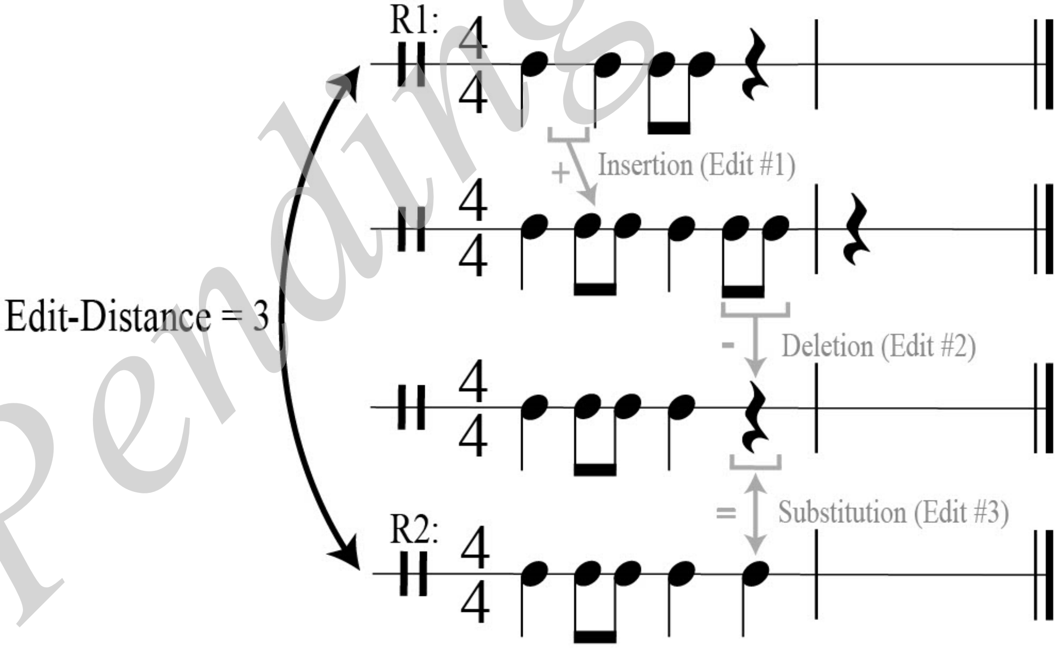

Figure 2. An example of an edit-distance of 3 between two rhythms (R1 and R2), calculated through insertions, deletions, and substitutions.

Tempo is a visceral characteristic that strongly influences the identity of songs (Cupchik, Rickert, & Mendelson, 1982; Gabrielsson, 1973). Specifically in electronic dance music (EDM), tempo is a primary dimension for classifying EDM subgenres and strongly influences perceived similarity of rhythms (Caparrini, Arroyo, Pérez-Molina, & Sánchez-Hernández, 2020; Honingh et al., 2015). Moreover, musical phrases have been conventionally mapped into discrete categories based upon tempo (e.g., slow vs. fast beats, or *adiago* vs. *allegro*) (Gabrielsson, 1973) presumably due to perceptual ease. Significant changes in tempo can inhibit the ability to recognize melodies (Halpern & Müllensiefen, 2008). For example, many musical genres and folk tunes are easily recognizable and discriminated based on tempo (Cupchik et al., 1982; Halpern, 1988), and dramatically sped up or slowed down versions of songs appear to change their identity. Additionally, fluctuations in tempo appear to alter the relative subdivision patterns and durations of individual notes within isochronous rhythms such as the samba, owing to the inextricable relationship between tempo and rhythmic content (Haugen & Danielsen, 2020). As such, tempo is an important factor to be included when evaluating the edit-distance model.

Overall, the present study sought to further augment the previous groundwork regarding edit-distance in rhythm similarity (Toussaint et al., 2012; Toussaint & Oh, 2016). We constructed a total of 16 rhythm phrases that independently varied in tempo and rhythmic structure with a few important constraints regarding the edit-distance manipulation (Figure 3). Although edit-distance encompasses three types of edits (substitution, insertion, and deletion), it is important to note that insertions and deletions add or remove a single rhythm unit, thereby altering the perceived meter of a rhythm phrase (Toussaint et al., 2012). As such, insertions and deletions can be more problematic when comparing rhythm phrases with an odd number of edits (e.g., 1, 3, 5, etc.), as this can change the meter of a rhythm phrase between duple and triple. By contrast, substitutions allow us to manipulate edit-distance while keeping meter constant (Toussaint et al., 2012). To best control for the potential confounding influence of metric changes (Cao et al., 2014; Prince, 2014), we limited our transformations of rhythm phrases to substitutions of individual rhythm units (i.e., sounded onsets of rhythm notation). Additionally, we substituted rhythm units that matched in total duration (e.g., quarter note and eighth note pairs) (Figure 3).

Each of the 8 unique rhythm phrases used in this study was generated at two different tempos—a moderate tempo of 90 beats per minute (BPM) and a fast tempo of 150 BPM—leading to 16 rhythm phrases total. These largely different tempos were chosen as opposed to two similar tempos, such as 110 BPM and 120 BPM, to ensure that participants could clearly perceive the tempo differences during the task. During the study, each rhythm stimulus was paired with one another and presented to participants sequentially, who then rated the perceived similarity of the two rhythms. We hypothesized that rhythms presented at the same tempo would yield higher similarity ratings than rhythms at different tempos, and we also predicted that edit-distance would reliably account for similarity ratings regardless of differences in tempo.

3. Methods

3.1 Participants

Nineteen participants (10 females; range = 18-27 years, *M* = 21.7 years; *SD* = 2.5 years) were recruited from The Ohio State University community. All participants gave written, informed consent approved by The Ohio State University Institutional Review Board. Data from one participant (1 female) was discarded due to an error in the experiment code, leaving a total of 18 intact participants' data. Before the experiment, participants filled out a survey about their demographic and musical background. Each participant's musical experience was quantified as the sum of the total number of years of formal experience including private lessons and class instruction. If participants played multiple instruments and/or had overlap in years of experience, then the overlapping years were counted only once. Overall, our participants had moderate musical experience (*M* = 5.7 years; *SD* = 5.6 years), but most were not currently engaged in any type of musical activities. Each participant received either monetary compensation or extra credit in a course for their participation.

3.2 Stimuli and Materials

Rhythm stimuli were created in *MuseScore* (version 2.1.0) as .wav files with a sampling rate of 44.1 kHz. All stimuli were created using the wood block instrument without any added reverb. Figure 3 shows eight rhythm phrases used in this experiment (referred to as R1 through R8), whose pairwise edit-distance was systematically varied from 1 to 4 solely through substitutions (Table 1). As an example, to derive R2 from R1 one would substitute the first quarter note of R1 with two eighth notes. Since one substitution was required, this demonstrates that R1 and R2 had a pairwise edit-distance of 1. Each of the 8 rhythm phrases was generated at two different tempos, once at quarter note = 90 BPM (beat period = 667ms) and again at 150 BPM (beat period = 400ms), yielding a total of 16 rhythm stimuli.

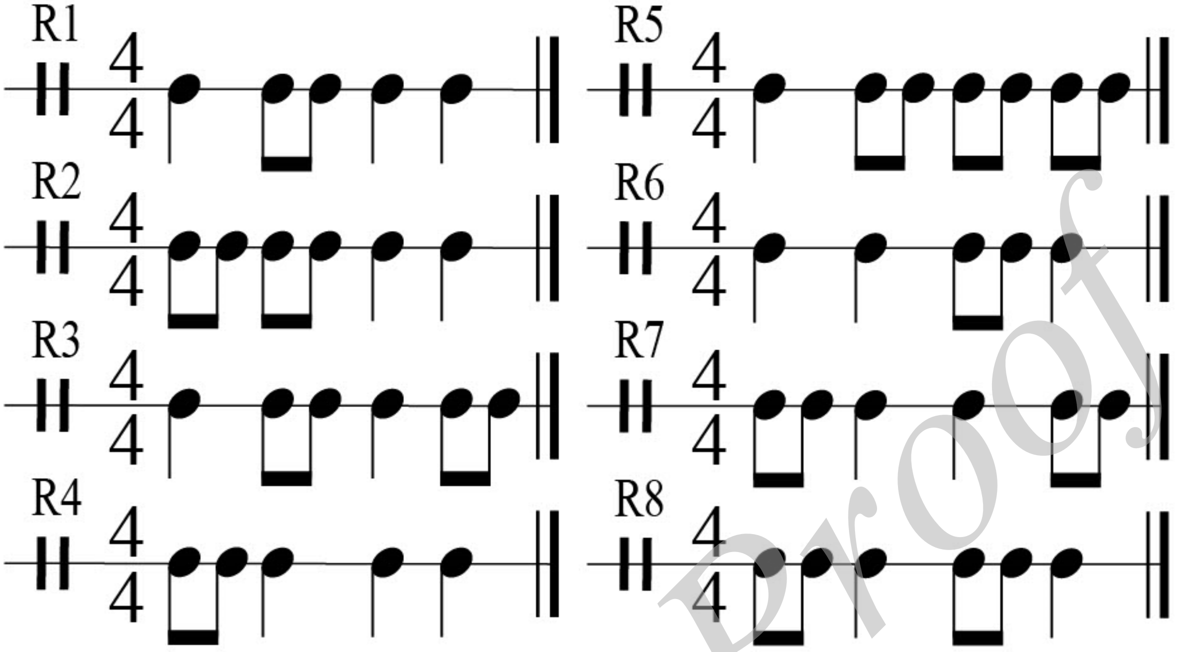

Figure 3. The stimuli used in the present experiment. The rhythm phrases (R1 through R8) were constructed at two tempos—90 and 150 BPM.

	Pairwise Edit-Distance							
	R1	R2	R3	R4	R5	R6	R7	R8
R1	0							
R2	1	0						
R3	1	2	0					
R4	2	1	3	0				
R5	2	3	1	4	0			
R6	2	3	3	2	2	0		
R7	3	2	2	1	3	3	0	
R8	3	2	4	1	3	1	2	0

Table 1. Theoretical edit-distance between each rhythm phrase (R1-R8).

3.3 Task and Procedure

The experiment was administered using *MATLAB* (version R2017a, MathWorks) and *Psychtoolbox-3* (version 3.0.14, Kleiner et al., 2007) in a sound-proof audio booth. Participants first read the experiment's instructions on the computer at their own pace, which read that they would be listening to pairs of “sound bites” and rating their similarity. Immediately following instructions, five practice trials were presented prior to the experimental trials to acclimate the participant to the task. These practice trials were excluded from analysis. Each trial started with the participant listening to a pair of rhythms, with a 2,500ms period of silence between the stimuli. Then, participants rated the rhythms’ similarity on a Likert scale from “1” (most different) to “4” (most similar) using a keyboard. Although this range coincided with the edit-distance manipulation, this was not intended to reflect one-to-one correspondence between the two scales. For every trial, participants were instructed to respond as quickly as possible within five seconds after the second rhythm ended. There was a burst of white noise immediately after each response, which served to indicate the end of the current trial; the white noise was also intended to discourage carry-over memory of the previous rhythm phrases. No training or feedback was provided on how to judge and rate similarity, and there were no hints about the edit-distance and tempo manipulations before the experiment.

Each of the sixteen stimuli were presented in all possible pairs within (e.g., 90 vs. 90 BPM or 150 BPM vs. 150 BPM) and between tempos (e.g., 90 vs. 150 BPM), including all 16 pairs of identical stimuli, resulting in a total of 136 trials (calculated as n(n+1)/2; where ‘n’ is the total number of stimuli). These were randomly presented across 4 blocks of 34 trials each. A self-paced recess occurred halfway into each block, and two minutes of mandatory recess occurred at the end of each block. In total, the task took approximately 25 to 30 minutes to complete.

3.4 Analysis

3.4.1 Inter-Participant Reliability

We first assessed how consistent similarity ratings among rhythm pairs were between participants. For each participant, similarity ratings of rhythm pairs were arranged into a distance (i.e., similarity) matrix. A test of congruence among distance matrices (CADM; Legendre & Lapointe, 2004) was used to evaluate the inter-participant agreement of similarity matrices. The CADM method tests the significance of Kendall's coefficient of concordance (Kendall's *W*) between multiple distance matrices. Kendall's *W* is a metric used to evaluate the rating agreement between participants, ranging from 0 (no agreement) to 1 (unanimous). This analysis creates a null distribution by repeatedly permuting the rows and the corresponding columns of each distance matrix and calculating Kendall's *W* from the permuted matrices. The significance of the observed coefficient is evaluated against the null distribution generated by permutation (n = 10,000). A strength of the CADM test is it allows for post hoc tests of whether and to what extent each participant's distance matrix is congruent with the others. Thus, the group-level CADM analysis was followed by *a posteriori* tests to further identify participants with deviating ratings. Analyses were implemented using the CADM package (Campbell, Legendre, & Lapointe, 2011) in *R* software (version 3.4.2).

3.4.2 Non-Metric Multidimensional Scaling

We employed non-metric multidimensional scaling (nMDS) in order to visualize participants' internal representation of the rhythm stimuli. Furthermore, the resulting dimensions of nMDS will be used in subsequent Mantel tests to scrutinize the edit-distance effect. Previously, metric MDS has been used to spatially map the perceptual similarity between musical stimuli based on categories including genre, tempo, and emotional valence (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005; Georges & Nguyen, 2019; Novello, McKinney, & Kohlrausch, 2006). One important advantage of nMDS over MDS in measuring perceptual similarity data is that it yields more consistent similarity distances among the items using the ordinal rank obtained from each participant whose extent of rating may considerably vary (Agarwal et al., 2007).

Individual similarity matrices were averaged into a group similarity matrix due to high concordance across participants (see Results). The average similarity matrix was used as input for nMDS in *R* software (version 3.4.2) using *RStudio* (version 1.1.383). Furthermore, the goodness of fit of the nMDS model is depicted by a quantity called 'stress' with 0 being most optimal (Kruskal, 1964). As such, we performed nMDS iteratively until the stress value fell below the acceptable limit (stress < 0.1) for optimal model fit (Novello et al., 2006).

3.4.3 Evaluation of Edit-Distance

To evaluate the edit-distance model, we separately created two similarity matrices containing mean ratings for within- and between-tempo conditions for each participant (2 per participant, 36 matrices total). Then, these individual similarity matrices were averaged to form a group-level similarity matrix per each condition. Finally, the two group-level matrices (Tables 2 and 3) were compared against the theoretical edit-distance matrix (Table 1) using the Mantel test, a non-parametric test of correlation between distance matrices. This analysis creates a sampling distribution by repeatedly permuting the rows and the corresponding columns of one matrix and calculating Spearman's correlation coefficients (Mantel, 1967; Legendre, 2000). The p-value is computed by comparing the data against a null distribution generated by permutation (n = 10,000). Each step of the Mantel tests was implemented using the *ncf* package in *R* software (version 3.4.2).

4. Results

4.1 Inter-Participant Reliability

The CADM test revealed a significant agreement of similarity ratings between participants ($W = .333$, $p < .0001$). A subsequent post-hoc congruence test further confirmed that every participant's ratings was consistent with the others (all $p < .001$). Although not every identical rhythm pair (e.g., the diagonal elements of Table 2) was rated as most similar with a "4.0" rating despite their exact same rhythmic content and tempo, the high concordance of ratings across participants and large majority of identical rhythms rated with the highest similarity rating (266 out of 288 trials) indicated that only a few participants experienced momentary and occasional lapses of attention during the experiment. Overall, these results assured reliable responses across all listeners, which were subsequently used in the nMDS and Mantel Test analyses.

4.2 Non-Metric Multidimensional Scaling

Optimal nMDS generated a total of 7-dimensional space (stress = 0.00613) when the stress value fell below the acceptable threshold (stress < 0.1). Among the 7 dimensions, only the first two dimensions were interpretable and no logical labels could be assigned to the rest (potential candidates for dimension labels included number and location of quarter and eighth notes). As shown in Figure 4, the first dimension (horizontal) clearly corresponded to the tempo of stimuli; rhythms at 90 BPM were clustered on the left side and rhythms at 150 BPM were clustered on the right side. The second dimension (vertical) of the nMDS map appeared to correspond to rhythm primacy—whether rhythm phrases began with a quarter note (the top half) or an eighth note pair (the bottom half). Note that rhythm primacy is not independent from edit-distance; shared primacy between two rhythms means that the maximum edit-distance between the rhythms is reduced by one. As such, the potential confounding effect of rhythm primacy on edit-distance will be considered in the following analysis of edit-distance. Together, nMDS analysis confirmed that the manipulation of tempo was successful, but it also newly yielded primacy as another important factor for rhythm similarity.

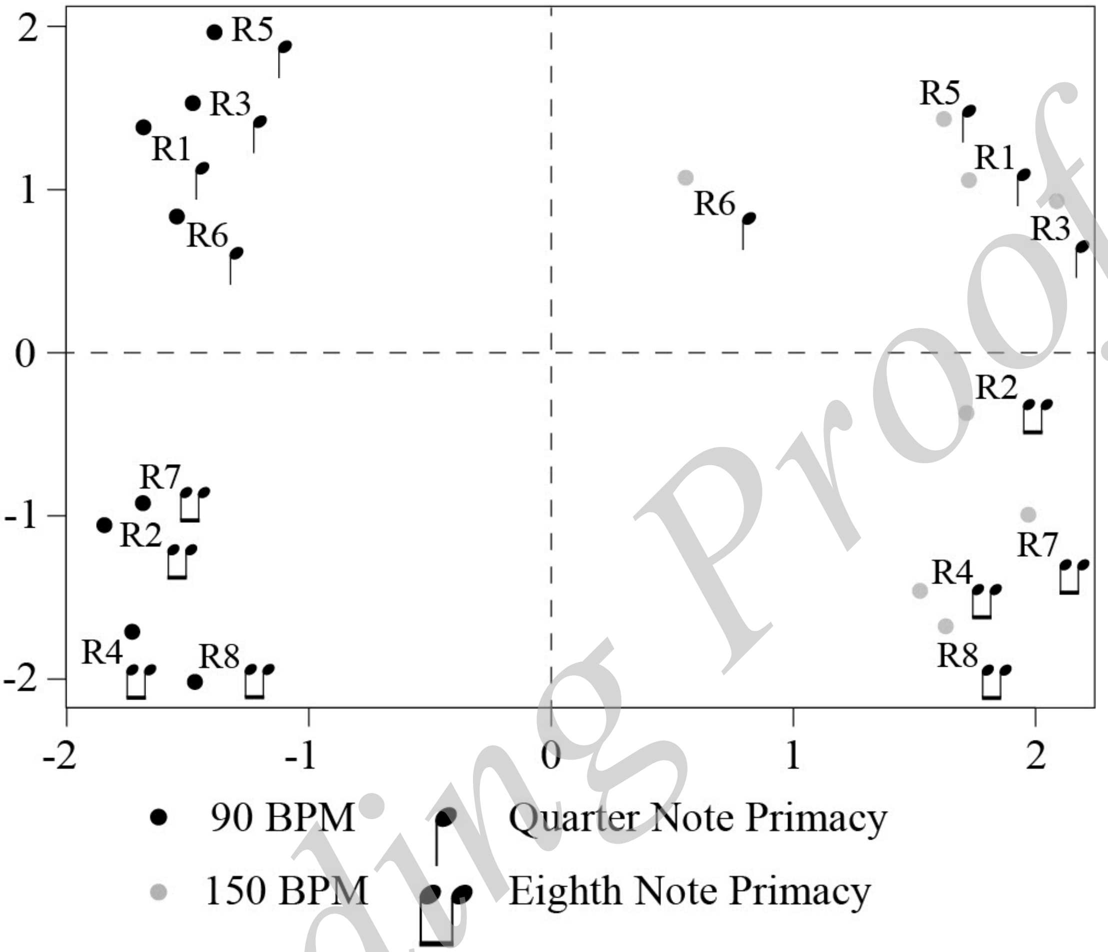

Figure 4. nMDS map visualizing the two distinct patterns of rhythm clusters. The horizontal dimension represents tempo, since rhythms clustered on the left side have a slower tempo of 90 BPM while rhythms on the right side have a faster tempo of 150 BPM. The vertical dimension represents rhythm primacy, with rhythms on the top half beginning with a quarter note and rhythms on the bottom half beginning with a pair of eighth notes.

4.3 Evaluation of Edit-Distance

The group-averaged similarity rating matrices for the within- and between-tempo conditions are shown in Tables 2 and 3, respectively. In line with the nMDS results, similarity ratings for the within-tempo rhythm pairs were overall higher than those in the between-tempo condition.

	Mean Ratings of Similarity (Within-Tempo)							
	R1	R2	R3	R4	R5	R6	R7	R8
R1	3.8							
R2	2.5	3.9						
R3	3.3	2.3	3.9					
R4	2.5	3.0	2.1	3.9				
R5	2.7	2.4	2.8	2.1	3.9			
R6	2.5	2.1	2.4	2.1	2.0	3.9		
R7	2.4	2.8	3.0	3.2	2.2	1.8	4.0	
R8	1.8	2.6	2.2	2.9	1.9	2.0	2.8	4.0

Table 2. Group-averaged ratings of similarity for each pair of rhythms in the within-tempo condition. Scores closer to 4 indicated “most similar” while closer to 1 indicated “most different.”

	Mean Ratings of Similarity (Between-Tempo)							
	R1	R2	R3	R4	R5	R6	R7	R8
R1	2.4							
R2	1.9	2.4						
R3	2.1	1.8	2.3					
R4	1.8	2.1	1.8	2.7				
R5	1.7	1.6	2.1	1.5	2.6			
R6	1.8	1.6	1.6	1.7	1.8	2.8		
R7	1.8	1.9	1.9	2.1	1.8	1.7	2.4	
R8	1.4	1.5	1.6	2.0	1.4	1.7	2.0	2.8

Table 3. Group-averaged ratings of similarity for each pair of rhythms in the between-tempo condition. Scores closer to 4 indicated “most similar” while closer to 1 indicated “most different.”

The effect of edit-distance on rhythm similarity was examined using Mantel tests on both within-tempo (Table 2) and between-tempo (Table 3) conditions by comparing the observed similarity matrices to the theoretical edit-distance matrix (Table 1). The tests revealed that the similarity ratings during the within-tempo condition were significantly correlated with edit-distance (r = –.648, p < .001), replicating previous findings (Toussaint & Oh, 2016; Toussaint et al., 2012). Moreover, edit-distance had a significant correlation with similarity ratings during the between-tempo condition (r = –.760, p <.001), indicating that edit-distance impacted rhythm similarity judgements even when the two rhythm phrases differed considerably in tempo. Figure

5 illustrates the correlations between the off-diagonal elements of the edit-distance matrix and the two similarity matrices.

Given that rhythm similarity was also influenced by primacy in the nMDS, we created a primacy distance matrix for use with the Mantel test in order to examine whether the effect of primacy was significant on similarity data. This primacy distance matrix had binary coding (1 or 0) differentiating whether rhythms had same or different beginning patterns. The Mantel test showed that the primacy matrix was significantly correlated with both similarity matrices for the within-tempo (*r* = –.645, *p* < .05) and the between-tempo (*r* = –.534, *p* < .05) conditions, which prompted us to examine whether the effect of edit-distance would be moderated by rhythm primacy for both within- and between-tempo conditions (Smouse et al., 1986). We performed the Mantel tests again with rhythm primacy being controlled, which revealed that the correlation between edit-distance and rhythm similarity ratings remained significant for both within-tempo (*r* = –.475, *p* < .01) and between-tempo (*r* = –.666, *p* < .001) conditions.

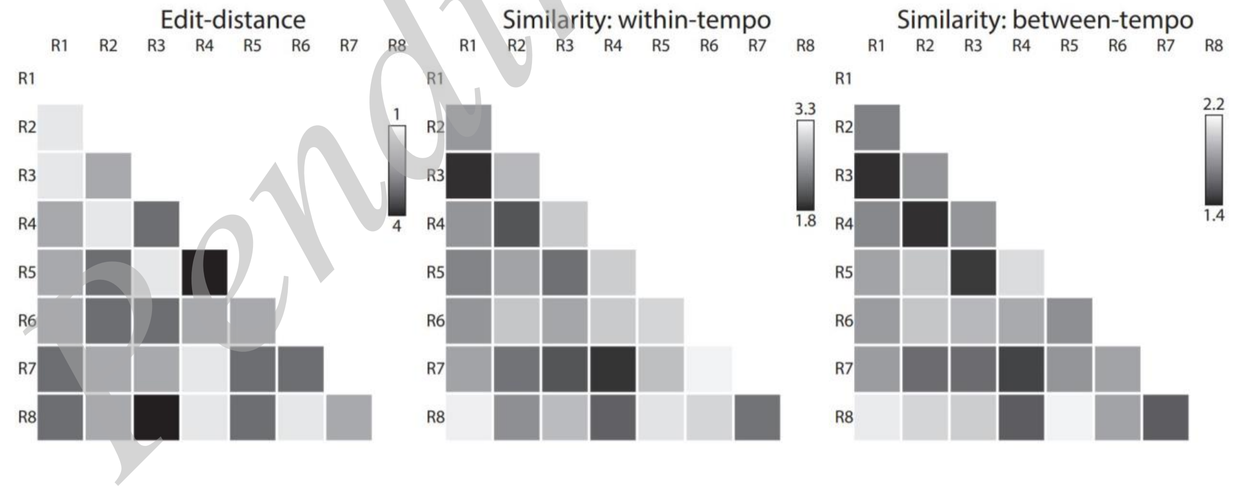

Figure 5. A visualization of the correlations between the off-diagonal elements of the theoretical edit-distance matrix (left) and the two group-level similarity matrices (middle and right). Lighter shades of grey represent rhythm pairs with either a smaller edit-distance or higher similarity.

5. Discussion

In the present study, we investigated rhythm similarity using the edit-distance model (Post & Toussaint, 2011; Toussaint & Oh, 2016; Toussaint et al., 2012). In particular, we were interested in whether or not edit-distance could account for the degree of perceptual similarity between unique rhythm phrases that also differed in tempo—a question hitherto unexplored despite its ecological importance. As expected, the nMDS visualized a robust clustering of rhythms on the basis of tempo, but the data-driven approach newly found that rhythms were also clustered on the basis of the onset pattern, a phenomenon we termed rhythm primacy. Mantel tests revealed that substitution-based edit-distance reliably accounted for perceptual similarity of rhythms irrespective of tempo. Finally, a partial Mantel test further confirmed the edit-distance effect while controlling for the effect of primacy.

Together, our findings lend further support to the edit-distance model (Toussaint et al., 2012; Toussaint & Oh, 2016). More importantly, we demonstrate for the first time that the edit-distance model can explain perceptual similarity across rhythmic phrases with different tempos. This is a crucial extension of previous literature, which only utilized rhythm phrases at the same tempo, raising a question of its ecological validity (Post & Toussaint, 2011; Toussaint & Oh, 2016; Toussaint et al., 2012). Natural music is multifaceted and contains wide variations in tempo, even within the same song, thus it can be challenging to develop algorithms that can accurately sort music that renders similar percepts. As such, our finding of tempo-invariant edit-distance offers further validation that edit-distance can also be an effective tool to help develop music classification algorithms (Esparza, Bello, & Humphrey, 2015; Lidy & Rauber, 2005; Meng Ahrendt, Larsen, & Hansen, 2007).

A fundamental question would be whether or not edit-distance is adopted as a plausible biological algorithm for rhythm analysis in music. None of the participants were able to consciously count the number of edits in order to transform one rhythm into another during the instantaneous response period after each trial. Nevertheless, participants' similarity ratings were remarkably in line with the theoretical edit-distance, and there was a robust consistency across participants' judgments. This suggests that analysis of edit-distance may be hard-wired in the human auditory system, which can immediately render perceptual gestalt of rhythmic patterns in music. Indeed, a recent fMRI study demonstrated that rhythmic gestalt was represented in the bilateral temporoparietal junction and right inferior frontal gyrus (Notter et al., 2019). In this study, a linear classification algorithm was used to probe every location of the brain that generated a spatially distributed pattern of neural activity across three short rhythm phrases collapsed across different tempos. However, it remains to be determined whether or not rhythms across different tempos elicit similar neural representations in these regions if their edit-distance is kept small.

When it comes to the perceptual gestalt of rhythms, tempo may provide the primary cue to discern the qualitative differences between rhythms. In the present experiment, listeners, with no hints, had to judge perceptual similarity of rhythmic pairs that spanned only one measure and were matched in other important musical characteristics such as timbre, pitch, and meter. Under such constraints, tempo provided listeners with an obvious criterion when discerning rhythm similarity, which was clearly visualized by the nMDS analysis. This is consistent with previous literature demonstrating that tempo differences influenced similarity ratings of existing music pieces (Cupchik et al., 1982; Honingh et al., 2015; but see also Novello et al., 2006). In other words, different songs with similar tempos were rated as more similar than different songs with markedly different tempos. In essence, tempo is intrinsic to rhythm similarity and is a dominant factor when judging perceptual similarity across different rhythmic patterns.

Furthermore, in the present study we opted to employ substantially different tempos (90 BPM vs. 150 BPM) for the rhythm stimuli for the purpose of ensuring that listeners were readily able to perceive the difference in tempo. However, this may have created unexpected interactions between the onset pattern (eighth vs. quarter note) and tempo. For example, a particular rhythm beginning with two quarter notes at 150BPM can be perceptually equivalent to another rhythm beginning with two eighth notes at 75 BPM. This was indeed the case, wherein one of the 150 BPM rhythms that began with two quarter notes (R6) was clustered closer with the 90 BPM rhythms in the nMDS.

Another unexpected finding from the nMDS analysis was a primacy effect in the absence of a recency effect. Typically, both primacy and recency effects are found in serial recall tasks (Greene & Samuel, 1986; Murdock, 1962; Roberts, 1986; Tzeng, 1973), but primacy effects are also often found in recognition tasks that are akin to the similarity judgment task employed in the current study (Digirolamo & Hintzman, 1997). Our finding of an isolated primacy effect may also be explained by the metrical organization of the rhythm stimuli. For example, beats 1 and 3 are strong in musical rhythms while beats 2 and 4 are weak in 4/4 meter (Lerdahl & Jackendoff, 1983; Phillips-Silver & Trainor, 2005). This metrical interaction could explain why the 1st beat (i.e., primacy) is more salient to listeners than the 4th beat (i.e., recency) (Jones, 2004).

Arguably, participants may have perceived some rhythm phrases as starting with an upbeat (i.e., anacrusis) instead of a downbeat, further impacting the primacy effect and rhythm similarity judgements. For example, rhythms beginning with two eighth notes (e.g., R4) could be interpreted as starting on an upbeat with a perceived (but not presented) stress on beat 2. Conversely, rhythms beginning with a quarter note (e.g., R1) could be perceived as starting on the downbeat. However, both meter and the number of beats were controlled to maintain a uniform structure for the rhythm phrases, and equal stress was placed on each of the four beats. Thus, the beat and meter were presented consistently across all subjects. Since the first onset of a rhythm phrase generally has the highest perceptual salience (Ladinig et al., 2009; Toussaint et al., 2012), it is unlikely that participants perceived our rhythm phrases as beginning with an anacrusis especially given the equal stress placed on each beat. Although melodies using short-short-long (SSL) rhythms could be perceived as starting on an upbeat compared to long-short-short (LSS) rhythms (London, Cross, & Himberg, 2006), this effect is not always consistent across individuals and different rhythmic structures (Stobart & Cross, 2008; Vos, Dijk, & Schomaker 1994). As such, we believe that the perception of anacruses, if present, is rather unpredictable and would not greatly impact our results.

The dominant influence of both primacy and tempo on the nMDS map raises the question of which factor could take priority when judging rhythm similarity. In real music, various genres of ballroom dances and EDM are defined by their restrictive tempo ranges (Dixon, Gouyan, & Widmer, 2004; Panteli et al., 2014). Since tempos are very similar between songs in these genres, judging rhythm similarity may rely on alternative factors than tempo to help compare the phrases, such as primacy in successive motifs. Furthermore, previous work has shown that ratings of rhythm similarity also appear to be influenced by the swing and metrical "feel" of a piece, a participant's musical experience, and the presence of musical context; rhythms heard as isolated phrases tend to be rated as more similar than when they are presented within the original piece of music (Bruford, McDonald, Barthet, & Sandler, 2019; Cameron, Potter, Wiggins, & Pearce, 2017). As such, the interactions between primacy and other external factors, such as musical experience and context, should be further surveyed.

Despite the presence of tempo and primacy which increases the ecological validity when evaluating the edit-distance effect, we note that our conclusions drew from a rather constrained set of rhythmic structures. For instance, we only opted to choose substitutions as a way of varying the edit-distance among rhythms although previous studies also included insertions and deletions (Toussaint et al., 2012; Toussaint & Oh, 2016). It is important to point out that we intended to keep the meter unchanged (Cao, Lotstein, & Johnson-Laird, 2014; Prince, 2014) while systematically varying the edit-distance, for which insertions and deletions were not viable options.

The present study also used limited stratifications of tempo and edit-distance. In our rhythm phrases, we set the maximum edit-distance at 4 (Figure 3) which is identical to the manipulation used in previous literature (Toussaint et al. 2012). However, trends in rhythm similarity ratings may be affected by a larger edit-distance range (e.g., edit-distance = 0-8), a larger range of tempos (e.g., 60 BPM, 120 BPM, and 180 BPM), or a smaller increment between tempos (e.g., 100 BPM, 120 BPM, 140 BPM). Rhythm phrases with a longer duration (i.e., two-measure phrases) allow for rhythms with a higher number of edits, allowing for more complex changes between rhythms. Furthermore, substitutions can also be more complex than were explored in the present experiment. For example, by changing a set of eighth notes to a set of triplets the primary unit of subdivision is changed which may alter judgements of rhythm similarity. Moreover, the strength of primacy in the presence of other salient rhythmic features, such as accents, syncopations, and rests, is unknown and should be further investigated. Overall, combining these complex rhythms with insertions and deletions will help to determine the robustness of edit-distance, primacy, and tempo in contexts that more accurately reflect everyday music listening.

6. Conclusion

Using rhythm stimuli that differed in their tempo and content, our data corroborated the robust nature of edit-distance, indicating its significant influence on rhythm similarity ratings regardless of differences in tempo or rhythm primacy. While our evaluation offers a glimpse into rhythm similarity and perception, future study is warranted to generalize the present findings to more complex rhythms, additional tempos, and longer pieces of music.

7. Declaration of Conflicting Interests

The authors have no conflicts of interest to disclose regarding this study.

8. Funding

This research was supported by The Ohio State University College of Arts and Sciences Small Grant.

9. Acknowledgements

The authors would like to thank Jay Myung for his helpful expertise and insight during data analysis.

References

Agarwal, S., Wills, J., Cayton, L., Lanckriet, G., Kriegman, D., & Belongie, S. (2007). Generalized non-metric multidimensional scaling. In *Artificial Intelligence and Statistics* (pp. 11-18).

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*, *19*(8), 1113–1139. https://doi.org/10.1080/02699930500204250

Bruford, F., Barthet, M., McDonald, Sk., & Sandler, M. (2019). Modelling Musical Similarity for Drum Patterns: A Perceptual Evaluation. *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*, 131–138. https://doi.org/10.1145/3356590.3356611

Cambouropoulos, E., Crawford, T., & Iliopoulos, C. S. (2001). Pattern Processing in Melodic Sequences: Challenges, Caveats and Prospects. *Computers and the Humanities*, *35*(1), 9–21. https://doi.org/10.1023/A:1002646129893

Cameron, D., Potter, K., Wiggins, G., & Pearce, M. (2017). Perception of Rhythmic Similarity is Asymmetrical, and Is Influenced by Musical Training, Expressive Performance, and Musical Context. *Timing & Time Perception*, *5*(3–4), 211–227. https://doi.org/10.1163/22134468-00002085

Campbell, V., Legendre, P., & Lapointe, F.-J. (2011). The performance of the Congruence Among Distance Matrices (CADM) test in phylogenetic analysis. *BMC Evolutionary Biology*, *11*(1), 64. https://doi.org/10.1186/1471-2148-11-64

Cao, E., Lotstein, M., & Johnson-Laird, P. N. (2014). Similarity and Families of Musical Rhythms. *Music Perception: An Interdisciplinary Journal*, *31*(5), 444–469. https://doi.org/10.1525/mp.2014.31.5.444

Caparrini, A., Arroyo, J., Pérez-Molina, L., & Sánchez-Hernández, J. (2020). Automatic subgenre classification in an electronic dance music taxonomy. *Journal of New Music Research*, *49*(3), 269–284. https://doi.org/10.1080/09298215.2020.1761399

Cupchik, G. C., Rickert, M., & Mendelson, J. (1982). Similarity and preference judgments of musical stimuli. *Scandinavian Journal of Psychology*, *23*(1), 273–282. https://doi.org/10.1111/j.1467-9450.1982.tb00441.x

Digirolamo, G. J., & Hintzman, D. L. (1997). First impressions are lasting impressions: A primacy effect in memory for repetitions. *Psychonomic Bulletin & Review*, *4*(1), 121–124. https://doi.org/10.3758/BF03210784

Dixon, S., Gouyon, F., & Widmer, G. (2004). Towards Characterisation of Music via Rhythmic Patterns. In *ISMIR*.

Esparza, T. M., Bello, J. P., & Humphrey, E. J. (2015). From Genre Classification to Rhythm Similarity: Computational and Musicological Insights. *Journal of New Music Research*, *44*(1), 39–57. https://doi.org/10.1080/09298215.2014.929706

Gabrielsson, A. (1973). Similarity ratings and dimension analyses of auditory rhythm patterns. 1. *Scandinavian Journal of Psychology*, *14*(1), 138–160. https://doi.org/10.1111/j.14679450.1973.tb00105.x

Georges, P., & Nguyen, N. (2019). Visualizing music similarity: Clustering and mapping 500 classical music composers. *Scientometrics*, *120*(3), 975–1003. https://doi.org/10.1007/s11192-019-03166-0

Greene, R. L., & Samuel, A. G. (1986). Recency and suffix effects in serial recall of musical stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(4), 517–524. https://doi.org/10.1037/0278-7393.12.4.517

Halpern, A. R. (1988). Perceived and Imagined Tempos of Familiar Songs. *Music Perception: An Interdisciplinary Journal*, *6*(2), 193–202. https://doi.org/10.2307/40285425

Halpern, A. R., & Müllensiefen, D. (2008). Effects of timbre and tempo change on memory for music. *The Quarterly Journal of Experimental Psychology*, *61*(9), 1371–1384. https://doi.org/10.1080/17470210701508038

Haugen, M. R., & Danielsen, A. (2020). Effect of tempo on relative note durations in a performed samba groove. *Journal of New Music Research*, *0*(0), 1–13. https://doi.org/10.1080/09298215.2020.1767655

Honingh, A., Panteli, M., Brockmeier, T., Mejía, D. I. L., & Sadakata, M. (2015). Perception of Timbre and Rhythm Similarity in Electronic Dance Music. *Journal of New Music Research*, *44*(4), 373–390. https://doi.org/10.1080/09298215.2015.1107102

Jones, M. R. (2004). Attention and timing. In J. G. Neuhoff (Ed.), *Ecological psychoacoustics*, (pp. 49-85) New York: Academic Press.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1-14.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*(2), 115–129. https://doi.org/10.1007/BF02289694

Ladinig, O., Honing, H., Háádén, G., & Winkler, I. (2009). Probing Attentive and Preattentive Emergent Meter in Adult Listeners without Extensive Music Training. *Music Perception*, *26*(4), 377–386. https://doi.org/10.1525/mp.2009.26.4.377

Legendre, P. (2000). Comparison of permutation methods for the partial correlation and partial mantel tests. *Journal of Statistical Computation and Simulation*, *67*(1), 37–73. https://doi.org/10.1080/00949650008812035

Legendre, P., & Lapointe, F.-J. (2004). ASSESSING CONGRUENCEAMONG DISTANCE MATRICES: SINGLE-MALT SCOTCH WHISKIES REVISITED. *Australian New Zealand Journal of Statistics*, *46*(4), 615–629. https://doi.org/10.1111/j.1467842X.2004.00357.x

Lerdahl, F., & Jackendoff, R. (1983). An Overview of Hierarchical Structure in Music. *Music Perception: An Interdisciplinary Journal*, *1*(2), 229–252. https://doi.org/10.2307/40285257

Lidy, T., & Rauber, A. (2005). Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In *ISMIR* (pp. 34-41).

London, J., Cross, I., & Himberg, T. (2006). The effect of tempo on the perception of anacruses. *Proceedings of the 9th ICMPC*, 1641-1647.

Lowrance, R., & Wagner, R. A. (1975). An Extension of the String-to-String Correction Problem. *Journal of the ACM*, *22*(2), 177–183. https://doi.org/10.1145/321879.321880

Mantel, N. (1967). Ranking Procedures for Arbitrarily Restricted Observation. *Biometrics*, *23*(1), 65–78. JSTOR. https://doi.org/10.2307/2528282

Meng, A., Ahrendt, P., Larsen, J., & Hansen, L. K. (2007). Temporal Feature Integration for Music Genre Classification. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*(5), 1654–1664. https://doi.org/10.1109/TASL.2007.899293

Murdock Jr., B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482–488. https://doi.org/10.1037/h0045106

Novello, A., McKinney, M. F., & Kohlrausch, A. (2006). Perceptual evaluation of music similarity. In *ISMIR*, 246-249.

Notter, M. P., Hanke, M., Murray, M. M., & Geiser, E. (2019). Encoding of Auditory Temporal Gestalt in the Human Brain. *Cerebral Cortex*, *29*(2), 475–484. https://doi.org/10.1093/cercor/bhx328

Orpen, K. S., & Huron, D. (1992). Measurement of similarity in music: A quantitative approach for nonparametric representations. *Computers in Music Research*, *4*, 1-44. Retrieved from https://csml.som.ohio-state.edu/Huron/Publications/orpen.similarity.text.html

Panteli, M., Bogaards, N., & Honingh, A. K. (2014). Modeling Rhythm Similarity for Electronic Dance Music. In *ISMIR* (pp. 537-542).

Paulus, J., & Klapuri, A. (2002). Measuring the similarity of Rhythmic Patterns. In *ISMIR*.

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the Beat: Movement Influences Infant Rhythm Perception. *Science*, *308*(5727), 1430–1430. https://doi.org/10.1126/science.1110922

Post, O., & Toussaint, G.T. (2011). The edit distance as a measure of perceived rhythmic similarity. *Empirical Musicology Review*, 6(3), 164–179. https://doi.org/10.18061/1811/52811

Prince, J. B. (2014). Contributions of pitch contour, tonality, rhythm, and meter to melodic similarity. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(6), 2319–2337. https://doi.org/10.1037/a0038010

Roberts, L. A. (1986). Modality and suffix effects in memory for melodic and harmonic musical materials. *Cognitive Psychology*, *18*(2), 123–157. https://doi.org/10.1016/0010-0285(86)90010-1

Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. *Systematic Zoology*, *35*(4), 627–632. JSTOR. https://doi.org/10.2307/2413122

Stobart, H., & Cross, I. (2000). The Andean anacrusis? Rhythmic structure and perception in Easter songs of Northern Potosí, Bolivia. *British Journal of Ethnomusicology*, *9*(2), 63–92. https://doi.org/10.1080/09681220008567301

Toussaint, G. T., Matthews, L., Campbell, M. & Brown, N. (2012). Measuring Musical Rhythm Similarity: Transformation versus Feature-Based Methods. *Journal of Interdisciplinary Music Studies*, *6*, 23-53. https://doi.org/10.4407/jims.2012.12.002

Toussaint, G. T., & Oh, S. M. (2016). Measuring Musical Rhythm Similarity: Further Experiments with the Many-to-Many Minimum-Weight Matching Distance. *Journal of Computer and Communications*, *4*, 117-125. http://dx.doi.org/10.4236/jcc.2016.415011

Tversky, A. (1977). Features of Similarity. *Psychological Review*, *84*(12), 327-352. http://dx.doi.org/10.1037/0033-295X.84.4.327

Typke, R., Veltkamp, R. C., & Wiering, F. (2004). Searching notated polyphonic music using transportation distances. *Proceedings of the 12th Annual ACM International Conference on Multimedia - MULTIMEDIA '04*, 128. https://doi.org/10.1145/1027527.1027551

Tzeng, O. J. L. (1973). Positive recency effect in a delayed free recall. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 436–439. https://doi.org/10.1016/S0022-5371(73)80023-4

Vos, P. G., van Dijk, A., & Schomaker, L. (1994). Melodic Cues for Metre. *Perception*, *23*(8), 965–976. https://doi.org/10.1068/p230965

Wagner, R. A., & Fischer, M. J. (1974). The String-to-String Correction Problem. *Journal of the ACM*, *21*(1), 168–173. https://doi.org/10.1145/321796.321811