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). pp. 2319-2337.

DOI: http://dx.doi.org/10.1037/a0038010

Contributions of pitch contour, tonality, rhythm, and metre to melodic similarity

Jon B. Prince

Murdoch University, Perth, Australia

Corresponding Author: Jon Prince
Murdoch University
90 South Street
Murdoch, WA 6150
Australia
tel (+61) 8 9360 6670
fax (+61) 8 9360 6878
j.prince@murdoch.edu.au

Keywords:
Pitch
Time
Music
Similarity

Abstract

The identity of a given melody resides in its sequence of pitches and durations, both of which exhibit surface details as well as structural properties. For the purposes of this research, pitch contour (pattern of ups and downs) served as pitch surface information, and tonality (musical key) as pitch structure; in the temporal dimension, surface information was the ordinal duration ratios of adjacent notes (rhythm), and metre (beat, or pulse) comprised the structure. Manipulating factorially the preservation or alteration of all of these forms of information in 17 novel melodies (typifying Western music) enabled measuring their effect on perceived melodic similarity. In Experiment 1, participants (N = 34, varied musical training) rated the perceived similarity of melody pairs transposed to new starting pitches. Rhythm was the largest contributor to perceived similarity, then contour, metre, and tonality. Experiment 2 used the same melodies but varied the tempo within a pair, and added a prefix of three chords, which oriented the listener to the starting pitch and tempo before the melody began. Now contour was the strongest influence on similarity ratings, followed by tonality, and then rhythm; metre was not significant. Overall, surface features influenced perceived similarity more than structural, but both had observable effects. The primary theoretical advances in melodic similarity research are that (1) the relative emphasis on pitch and temporal factors is flexible, (2) pitch and time functioned independently when manipulated factorially, regardless of which dimension is more influential, and (3) interactions between surface and structural information were unreliable and never occurred between dimensions.

Introduction

An enduring question in human perception is what makes two melodies sound similar. In fact, music is an especially well-suited domain for examining the general concept of similarity, as it consists of clearly-delineated dimensions that not only exhibit hierarchical structure and statistical regularities, but can be manipulated independently while preserving the naturalistic properties of the stimulus. Although music has multiple dimensions, pitch and time have received the most attention – likely because for the overwhelming majority of music, they together define the identity of a musical piece and exhibit the greatest degree of complexity. This complexity makes it difficult to sort out the details of how all the components of pitch and time contribute to perceived similarity. Indeed, how pitch and time combine in music perception remains an open question (for reviews, see Prince, Thompson, & Schmuckler, 2009; Schellenberg, Stalinski, & Marks, 2013).

Pitch and time have critical information at both the superficial surface level and at deeper structural levels (Krumhansl, 2000), as explained below. The aim of this article is to examine how surface and structural information in both pitch and time affect perceived melodic similarity, and in particular, how they combine. For the purposes of this article, surface information refers to pitch contour and rhythm (explained below), as they are comprised of information directly available at the level of the musical surface[1]. The structural information in this case is tonality and metre (explained further below), as they represent information derived from the surface.

Pitch contour refers to the pattern of ascending and descending pitch intervals of a melody, and it is a primary component of melodic perception (for reviews, see Deutsch, 2013; Schmuckler, 2009). Dowling (1978) presents contour as one of two critical factors (tonality being the other) in melodic perception and memory, showing that a non-exact imitation of the standard melody is often confused as a match when it has a similar contour.

Indeed, even when wildly out of tune, singers preserve the general contour of a melody (Pfordresher & Mantell, 2014). The importance of contour is also evidenced by its early emergence – infants as young as 5 months differentiate melodies primarily on the basis of their contour (for a review, see Trehub & Hannon, 2006).

The other component to Dowling's model of melodic perception is based on tonality (musical key), which refers to the hierarchical organisation of the 12 unique pitch classes per octave used in Western music, arranged around a central reference pitch, or tonic. For instance, in the key of G major, the pitch class G is the tonic – it is the most psychologically stable pitch and central cognitive reference point; all other pitches are ordered in a hierarchical fashion relative to the tonic. Tonics are heard more frequently, make better endings for melodies, and confer processing benefits (Krumhansl, 1990). Tonality is a fundamental characteristic of music, functioning as a structure on which to encode additional information (Dowling, 1978) and therefore is a strong contributor to melodic processing (for a review, see Krumhansl & Cuddy, 2010). Although some methodologies show musicians as more sensitive to tonality than untrained listeners (Krumhansl & Shepard, 1979), tonality strongly influences music perception regardless of expertise (Bigand & Poulin-Charronnat, 2006), even for "tone-deaf" individuals (Tillmann, Gosselin, Bigand, & Peretz, 2012). Looking more generally than the music cognition literature, physics experts tend to emphasise structural information in problem categorisation at the expense of surface, relative to novices (Chi, Feltovich, & Glaser, 1981).

Pitch cannot function alone in music – it is structured in time. Patterns of duration and relative timing comprise the rhythm, or temporal surface information, which has a strong role in melodic processing (for a review, see McAuley, 2010). Although any rhythmic change will decrease melodic recognition, not all aspects are equally influential. For example, Schulkind (1999) found that preserving the relative pattern of short and long notes (i.e., rhythm) while changing their absolute ratios (e.g., changing .2s, .6s, .3s to .1s, .6s, .4s) impaired recognition less than reordering the original durations (e.g., .6s, .2s, .3s).

Repeating patterns in rhythmic sequences lead to the abstraction of an underlying metrical pulse (beat), or metre (Lerdahl & Jackendoff, 1983). This hierarchical temporal structure (Palmer & Krumhansl, 1990) guides our attention (Jones & Boltz, 1989), improves the processing of events that coincide with the pulse (Barnes & Jones, 2000), modulates our interpretation of ambiguous rhythmic sequences (Desain & Honing, 2003), and influences perceived melodic similarity (Eerola, Järvinen, Louhivuori, & Toiviainen, 2001). There is also a prodigious literature on the role of metre in sensorimotor synchronisation (for reviews, see Repp, 2005; Repp & Su, 2013).

To perceive a melody, the listener must integrate the surface and structural information in both pitch and time, but how this occurs is unclear, particularly with regard to independence or interaction. In the case of contour and tonality (both pitch variables), they are theoretically independent in that any number of different pitch sequences can establish a tonal centre (musical key). Of course, the exact choice of pitch classes will determine whether the sequence is tonal, but the general up-down *shape* of the pitch profile does not restrict its tonality.

Accordingly, the majority of experimental evidence suggests that contour and tonality are processed and function independently (Dowling, Kwak, & Andrews, 1995; Edworthy, 1985; Eiting, 1984; Trainor, McDonald, & Alain, 2002). Dowling and colleagues have established that when comparing novel melodies with no delay, listeners primarily rely on contour – they are likely to falsely recognise a melody in the same key as a match if it has a similar contour. But for longer delays with interspersed melodies, listeners abstract a more detailed representation of the melody that is key-invariant and is more sensitive to structural information (Dewitt & Crowder, 1986; Dowling, 1978; Dowling et al., 1995). This differential contribution of tonality and contour to melodic memory implies independence of function. Repeated listenings also result in more observable effects of structural features on melodic perception, such as tonality (Pollard-Gott, 1983; Serafine, Glassman, & Overbeeke, 1989). Further, when sequences are atonal (not conforming to any musical key), listeners primarily rely on contour for processing melodies (Freedman, 1999; Krumhansl, 1991), also consistent with independence.

However there is contrary evidence, such as findings that tonality only matters when the contour information is preserved – without a matching contour, violating tonality had no effect on melody recognition (Massaro, Kallman, & Kelly, 1980). Additionally, the exact arrangement of intervals in 3-note sequences can influence the ease of establishing tonality (Cuddy & Cohen, 1976). Interestingly, the reverse pattern has also been reported – where processing contour information is easier for tonal melodies (Bartlett & Dowling, 1988; Cuddy, Cohen, & Mewhort, 1981; Dowling, 1991). Thus tonality and contour may not be fully independent.

For rhythm and metre (both temporal variables), it is again the case that any number of different surface (rhythmic) patterns may instantiate a given structure (metre), suggesting some degree of theoretical independence. Although the particular sequence of time intervals between events determines whether a metrical framework can be extracted from a rhythmic pattern, the ordinal sequence itself does not necessarily constrain its potential metrical interpretations. However the exact sequence of intervals is not trivial – rhythmic patterns are a primary factor in establishing the perception of musical events, such that the occasional long gap between events in a sequence indicates a grouping boundary (Garner & Gottwald, 1968). Inter-onset intervals that are related with regular simple integer ratios (e.g., 1:2, 1:3) can go on to establish metrical frameworks (Povel & Essens, 1985), but even those with complex ratios (e.g., 1:2.5, 1:3.5) can successfully form into groups and be processed with (admittedly lower) accuracy (Essens, 1986; Essens & Povel, 1985; Handel & Oshinsky, 1981), as well as learned implicitly (Schultz, Stevens, Keller, & Tillmann, 2013). Thus the sequence of durations in a rhythmic pattern has unique importance beyond its role in establishing a metre (Monahan, Kendall, & Carterette, 1987). Nevertheless, it is unlikely that rhythm and metre can function entirely independently – not only is metre extracted from the rhythmic surface, but the metric framework can modify perception of rhythmic sequences (Desain & Honing, 2003).

What about cross-dimensional relations in surface and structure? For instance, can tonality affect rhythm perception, or metre affect contour perception? This question is even more difficult to answer, partly because the relation between pitch and timing information varies greatly depending on the stimuli and task (Barnes & Johnston, 2010; Prince, 2011; Tillmann & Lebrun-Guillaud, 2006). Melodic recognition accuracy decreases when the standard and comparison melodies have different rhythmic groupings (Dowling, 1973; Jones & Ralston, 1991) or metrical frameworks (Acevedo, Temperley, & Pfordresher, 2014). Increasing the tempo of interleaved melodies fosters their segregation into separate streams, although this requires alternations between low and high pitches at extremely rapid rates of less than 150 ms between tones (Bregman, 1990). The exact combination of rhythmic and melodic patterns can also influence the ability to discriminate targets and decoys (Jones, Summerell, & Marshburn, 1987), although in that study listeners only used rhythmic patterns to differentiate melodies if the decoy contour remained the same. Boltz (2011) found that raising the pitch or brightening the timbre of melodies makes them seem faster. Using trained musicians only, Abe and Okada (2004) reported that shifting the phase of pitch and temporal patterns (by 1-2 positions) altered the interpretation of the musical key, but not the perceived metre, thus an asymmetric relationship between metre and tonality. However, other research found the opposite asymmetry, in which musicians were more likely to report that probes following a melody were on the beat if its pitch was tonally stable, but pitch judgements were unaffected by their metrical position (Prince, Thompson, et al., 2009).

Research on the relative contribution of pitch and time to perceived similarity of novel melodies generally finds that temporal surface information is most prominent. Halpern (1984; Halpern, Bartlett, & Dowling, 1998) analysed the similarity ratings of 16 melodic sequences, and found that changes to the rhythmic properties were most influential on ratings, followed by contour, and then whether the melody was in a major or minor key (tonal structure). Rosner and Meyer (1986) also reported that rhythm (the temporal surface) was the most important factor on similarity of 12 melodies, followed by a mixture of surface and structural pitch variables.

Using qualitative descriptions of nine extracts from two musical pieces, Lamont and Dibben (2001) highlighted the role of surface features such as dynamics (loudness) and tempo over pitch height and contour. Moreover, these authors found no role of deeper structural information. McAdams, Vieillard, Houix, and Reynolds (2004) asked listeners to group 34 sections of a single musical piece according to their own subjective criteria (i.e., no predefined categories), and then provide terms that capture the essence of what makes a group similar. Temporal surface (tempo and rhythm) descriptors were the most prevalent and dominant characteristics, over pitch surface variables (average pitch height, contour).

Eerola et al. (2001) predicted the perceived musical similarity of 15 folk melodies based on statistical properties (frequency-based surface information) and descriptive characteristics (akin to structural information). The descriptive variables accounted for more variance than the statistical ones, but the best solution came from a combination of both variable types. They acknowledged the possibility of overfitting the data, but it is nonetheless important that both forms of information can contribute uniquely to perceived similarity.

None of the studies mentioned above directly address the relationship *between* pitch and temporal information in melodic similarity beyond their relative contribution – that is, how might they affect one another? In fact, there is only one article that touches on this issue (Monahan & Carterette, 1985). These authors found that five dimensions best explained similarity ratings of 32 melodies; the first three reflected temporal characteristics, and the last two were pitch-based. But the most immediately relevant result to cross-dimensional relations was an individual differences tradeoff between reliance on pitch and temporal information – participants who placed strong weight on temporal factors de-emphasised the pitch factors, and vice versa.

The question of interactions between the parameters of contour, rhythm, tonality, and metre requires a delicate balance between methodical experimental control and natural musical context. Because listeners in McAdams' et al. (2004) study established their own subjective criteria, it is difficult to establish quantitative interpretations of the data, and moreover, the attributes covaried – as would be expected in normal music heard in more naturalistic conditions. Rosner and Meyer (1986) stated that there should be interactions between them, but were not able to directly assess this possibility. Similarly, when explaining the relative lack of explanatory value of some of their measured variables, Eerola et al. (2001) pointed to the fact that their melodies varied simultaneously on multiple dimensions, and they were using an "oversimplified representation" of the melodies in their analyses. They recommended that future research vary the stimuli in a more systematic and controlled manner to assess more exactly their relative contribution.

Experiment 1

As stated earlier, the main goal of the present research is to examine how both surface and structural information in the dimensions of both pitch and time combine in contributing to melodic similarity. On the basis of the background literature, it is proposed that (1) surface information should be more influential than structure for novel melodies, and (2) temporal manipulations should have greater effect than pitch. However the primary theoretical question is to test for interactions *between* these variables (contour, rhythm, tonality, and metre), not only their respective roles. Because the background literature provides no clear guidance on this issue, the present research approaches this issue methodically by using a factorial manipulation of all these variables. Additionally, a much larger stimulus set than typically employed was created, using 17 typical melodies as starting points for creating 16 variants that factorially preserved/destroyed the contour, rhythm, tonality, and metre of the original melody (giving 272 unique sequences). Accordingly, no listener heard a given sequence twice, greatly reducing the potential role of learning during the experimental session affecting similarity judgements. Three analysis techniques were employed, including categorical ANOVA analyses (made possible by the factorial design), linear regression with non-intercorrelated predictors, and factor analysis. Together, this approach is intended to provide a close quantitative examination of the roles of contour, rhythm, tonality, metre in melodic similarity, and in particular how they combine.

Method

Participants. There were 34 participants, with an average age of 22.6 (*SD* = 4.7), and 3.5 years of musical training (*SD* = 4.8). Participants were recruited from the Murdoch University community, largely undergraduate psychology students. Compensation was either course credit or $10.

Stimuli. There were 17 normal melodies (*M* length = 12.1 notes, 4.8 seconds) that served as original seed melodies from which all 16 variants were created. The seed melodies were all in “common time” (4 beats per measure); 12 used the major scale and 5 using the melodic minor scale. The pitch and temporal characteristics of the seed melodies were varied independently, in factorial fashion. Table 1 summarises the manipulation levels and their properties, which are explained in detail below. There were four levels of pitch manipulation, where the first level (p1) was the original pitch sequence, that is, unaltered from the original melody. The melodies strongly established a musical key, as assessed by the Krumhansl-Schmuckler keyfinding algorithm (Krumhansl, 1990; Krumhansl & Schmuckler, 1986) – the average correlation coefficient of the distribution of pitches in p1 sequences with the intended key was .84 (*SD* = .08). This coefficient is known as the maximum key correlation (MKC).

The p2 level preserved the global pitch contour of its corresponding seed melody, but had a different set of pitches in order to destroy the sense of musical key (i.e., they did not fit in any Western major or minor key). The artificial set of pitches (or scale) consisted of A B C# D D# F G; like other scales, it could be transposed to start on any pitch. This scale preserves important characteristics of musical scales (Trehub, Schellenberg, & Kamenetsky, 1999) in that it used 7 of 12 pitch classes per octave, neighbouring pitches were either 1 or 2 semitones apart (1 semitone is the smallest possible step in Western music), and not all steps were equally sized. The p2 level therefore corresponded to a preservation of surface (contour) but violation of structure (tonality). Comparing the contour of corresponding p1 and p2 sequences by converting their notes to a series of pitch heights (e.g., 1, 4, 3, 6 and 2, 4, 3, 5) and correlating them resulted in a high level of agreement (*M r* = .93, *SD* = .06). Conversely, the average MKC of p2 sequences was low (*M* MKC = .44, *SD* = .11) compared to the much higher average p1 MKC (see above). The first and last notes of the sequence were unchanged from the seed melody, which were also members of the artificial scale.

A contour-violated manipulation level (p3) pseudo-randomly shuffled the order of the seed melody pitches, but did not add or delete any pitches. The randomisation had the constraints that the first and last pitch had to stay the same as the seed melody, but no other note could remain in its original place. This change therefore retained the tonality (structure) of the seed melody (*M* MKC = .83, *SD* = .09), while disrupting its contour (surface), as the average correlation of p3 and p1 pitch sequences was *r* = .14 (*SD* = .34).

The final pitch manipulation level (p4) was a contour-violated-atonal variant created by pseudo-randomly shuffling the order of the atonal p2 level, thereby destroying both the surface and structure of the seed melody. The randomisation constraints were the same as those used for creating the p3 level from the p1 level, but instead were applied to the p2 level. The average MKC of the p4 sequences was .50 (*SD* = .12) and the average correlation of p3 and p4 pitch sequences was .04 (*SD* = .35).

The four levels of time manipulation were also factorial variations of surface and structure. There were no silent gaps between notes for all levels, so durations were equivalent to inter-onset intervals. The t1 level was the original sequence of durations, which for each seed melody had 4 unique duration values: 167 ms (eighth note), 333 ms (quarter note), 500 ms (dotted eighth note), or 667 ms (half note). All t1 levels had a regular beat and were clearly metric, as measured by comparing the distribution of note onsets with the idealised metric hierarchy of Palmer and Krumhansl (1990); the average correlation was .78 (*SD* = .05).

The t2 level was an ametric variant that preserved the rhythmic pattern of the seed melody. This manipulation was accomplished by changing each of the 4 regular note durations used in the t1 level to a matched nearby value (200, 280, 530, and 650 ms, respectively). These new durations preserved the surface pattern of relative short and long durations (i.e., rhythm), but destroyed the temporal structure (metre). Whereas the original durations are related by simple integer ratios (1:2, 1:3, 2:3) that establish a regular beat, the new durations used complex integer ratios (e.g., 5:7, 20:53, 4:13) that did not accommodate any regular metric framework, thus violating the temporal structure. The average correlation of the series of durations comprising the rhythm of t1 and t2 sequences was .98 (*SD* = .01), demonstrating excellent preservation of the temporal surface.

The t3 level pseudo-randomly shuffled the order of the seed melody durations, thus creating a rhythm-violated sequence that preserved the metrical structure of the melody, in that all durations still accommodated a regular metrical framework. The randomisation had the constraints that the first and last duration had to stay the same as the seed melody, but no other duration could remain in its original place. Retaining the same quantised durations was largely successful in preserving the metrical framework, although the randomisation of duration order did result in a weaker correlation with the Palmer and Krumhansl (1990) hierarchy (*M* = .63, *SD* = .14). The surface information was demonstrably altered, as the average t1-t3 duration sequence correlation was .16 (SD = .27).

The final time manipulation level (t4) violated both the rhythm and the metrical framework of the seed melody, by pseudo-randomly shuffling the order of the t2 durations, using the same constraints as those for generating the t3 level. The average correlation of t3 and t4 duration sequences was .12 (*SD* = .20).

Combining all 4 pitch levels with 4 time levels generated 16 variants of each of the 17 seed melodies (see Table 2). Figure 1 depicts some example variants from one given seed melody. In a given trial, participants heard two sequences, consisting of two variants of the same seed melody (e.g., p1t4 and p2t3), and judged their similarity. That is, both melodies in a trial were derived from the same seed melody, never different seeds. Regardless of pitch manipulation level, the second melody of a pair always started on a different pitch (i.e., transposed to a new key), in order to avoid a confound between the manipulations of interest and the number of pitches shared between sequences, which affects perceived similarity (van Egmond & Povel, 1996; van Egmond, Povel, & Maris, 1996). A tonal sequence and an atonal sequence must have mostly different pitch classes because the scale has changed. Thus comparisons between two tonal sequences should share a similar number of pitch classes as a tonal-atonal pair, in order to separate the effects of pitch class overlap from tonality on perceived similarity. Transposing the melodies to different keys met this need, providing a way to control the number of shared pitch classes between sequences, thus preventing a confound between the manipulations of structure and surface.

Four different starting pitches were used for tonal sequences (C, D, E, and G#) and a different four for atonal sequences (C, C#, D#, G#). The assignment of starting pitches was arranged such that the average number of shared pitch classes ranged between 3.4 and 4.3. Tempo remained constant throughout the experiment in order to control the effects of elapsed time between standard and comparison on the memory trace. Experiment 2 returns to this issue. Melodies were generated as MIDI files in MATLAB and then converted to .wav files, all using the same piano timbre soundfont.

Comparing the 16 different variants provided 16 x 16 = 256 possible combinations for each seed melody (counting both orders of a given pair). Having all participants rate each combination would have made the experimental session too long. Instead, participants only heard one order of each variant combination (e.g., p2t3-p1t1 *or* p1t1-p2t3), giving 136 trials including match conditions such as p3t4-p3t4. The session took an average of 31 minutes to complete. Order combination was counterbalanced, sampling equally from above and below the diagonal of the 16 x 16 matrix for each participant. Also counterbalanced across participant was the assignment of melodies and variants, such that each participant never heard the same variant of a given melody more than once. Although a given trial consisted of variants derived from the same seed melody, subsequent trials would be based on a different seed melody.

Procedure. Participants gave informed consent and completed a background questionnaire on musical experience. The experimenter explained the task of rating melodic similarity, and also the concept of transposition by explaining that singing Happy Birthday starting on a low note or a high note did not change the melody. That is, it was the *pattern* of pitches that was important, not the absolute frequencies themselves.

Each trial began by the participant pressing the space bar, after which the first sequence of the pair began. They then had to press the ‘s’ key to hear the second sequence. This procedure ensured that the participants were aware of the separation between sequences, and was intended to eliminate confusion about when the first sequence ended and second began. On average, participants waited 0.90 seconds (*SD* = .42, median = 0.77) between sequences. Immediately following the second sequence, participants were prompted to provide a rating of similarity on the scale of 1 (not at all similar) to 7 (very similar).

Participants completed 3 practice trials before beginning the full experiment. The first practice trial presented a p1t1-p1t1 combination – an exact transposition of an original seed melody. If they gave a similarity rating below 6 (suggesting confusion regarding the transposition of the second sequence), the experimenter explained that this case was indeed an exact match, that is, both sequences had the same pattern of intervals and durations despite starting on different pitches, and that this was as similar as the sequences could get. Further, the experimenter re-explained the concept of transposition to ensure that the participant understood the task fully. The remaining practice trials consisted of a p3t1-p3t4 and a p1t2-p2t2 pair (randomising both within-trial sequence order and between-trial pair order); no further instructions regarding a “correct” rating were provided.

Data analysis. Before the main rating data analyses, there were preliminary inspections comprised of manipulation checks, examination of effects of variant order combination, and testing for expertise effects. Subsequently, an ANOVA tested the role of change (i.e., same or different within a trial) of contour, rhythm, tonality, and metre in a categorical analysis (thus a 2 x 2 x 2 x 2 univariate equation), made possible by the factorial design of the experiment. This analysis collapsed across participant (after first ensuring decent inter-participant agreement) [2], and enabled systematic evaluation of the interaction or independence of all manipulated variables.

The second approach was a linear regression equation (following Eerola et al., 2001) predicting the perceived similarity ratings averaged across participant, using continuous objective measures of contour, rhythm, tonality, and metre. The contour predictor was the average correlation coefficient of the two melodies, when coded as a numerical series of pitches[3]; the rhythm predictor was the average correlation of the sequence of durations (ms). Higher coefficients indicate greater predicted surface similarity, thus positive correlations between these variables and similarity ratings were expected. The tonality predictor was the average absolute difference in tonality (MKC) between the two melodies; the metre predictor was the average absolute difference in the correlation with the metric hierarchy (Palmer & Krumhansl, 1990) between the two melodies. As larger tonal and/or metric difference should result in lower similarity ratings, negative correlations were expected between these predictors and similarity ratings.

The final analysis approach involved exploratory factor analysis using principal components analysis of the 16 x 16 matrix of perceived similarity ratings averaged across participant (following Monahan & Carterette, 1985). These techniques allow extraction of the underlying factors that explain the similarity ratings while making no assumptions about the nature of the stimuli or experimental manipulations (Kruskal & Wish, 1978). The extracted factors are then inspected for the extent to which they resemble the manipulated differences between the melodies.

Results

Preliminary checks. To see if participants were able to notice changes in melodic similarity (that is, that the task was not too difficult), the average ratings for the 16 exact match conditions (e.g., p2t3-p2t3) was compared to the average rating from all 240 non-match conditions. Regardless of participants' use of surface and structure information in pitch and time, they should rate exact match conditions as more similar than non-matches. Reassuringly, the average similarity rating for the match conditions was 5.71 (*SD* = .46), compared to 4.30 (*SD* = .81) for the non-matches, demonstrating that participants were indeed sensitive to alterations to the melodies, $t(33) = 16.6$, $p < .001$ (all t-tests are two-tailed paired samples). Note that some non-matches were relatively similar, such as p1t1-p1t2, so an average rating of 4.30 for all non-matches is not unreasonable – by comparison, the average p1t1-p4t4 rating was 2.72. Figure 2 shows a greyscale plot of the 16 x 16 similarity matrix, averaged across participant. This figure reflects the fact that perceived similarity between melody pairs is high along the ascending diagonal (match conditions) and decreases with surface and structural differences in both dimensions. Note also the uniformly low ratings of the descending diagonal (conditions in which both the pitch and time levels were maximally different).

Order effects. Figure 2 is also useful for assessing the possibility of order effects – that the similarity rating between two variant types (e.g., p1t2 and p4t3) varies based on which type occurred first. Cells below the ascending diagonal (lower triangle) represent conditions in which the variant with fewer changes to the original melody (e.g., p1t2) is heard first, whereas above the diagonal (upper triangle) shows the variant with more changes (e.g., p4t3) first. It is possible that hearing a more typical melody followed by a less typical one would result in lower similarity ratings than the other direction (Bartlett & Dowling, 1988) because "good patterns have few alternatives" (Garner, 1970). Indeed, in the present data the average similarity rating of the lower triangle conditions was significantly lower than those in the upper triangle, $t(33) = 4.1$, $p < .001$, although the mean difference between triangles was only 0.2 (lower $M = 4.2$, $SD = .61$; upper $M = 4.4$, $SD = .57$).

Given this overall mean difference, further examination tested if the ratings in the lower and upper triangles followed the same pattern – that is, if the contribution of pitch and time manipulations changed as a function of variant order. The selected approach was to compare the consistency among participants (i.e., random variation) to that between the lower and upper triangles (variant order). Put differently, was the variation in ratings based on variant order (upper or lower triangle) comparable to what one would predict based on random variation between participants? Each participant experienced one of two possible variant order combinations, as participants did not hear all 256 variant order combinations (cf. last paragraph of Stimuli in the Method section). Therefore rating consistency had to be calculated separately for the two variant order combinations, grouping together participants who experienced the same variant orders. To examine the random variation between participants, each group was split into two subgroups (random assignment), whose ratings were averaged separately and correlated (using only the 120 non-diagonal cells). Participant subgroups intercorrelated at $r(118) = .60$ (for variant order group 1) and .64 (for variant order group 2), both *p*s <.001. This measure of the random between-participant variation was comparable to the correlation of the lower and upper triangles (averaging across participant), $r(118) = .59$, $p < .001$. In other words, ratings varied as much between participants (of a given group) as they did across the overall lower/upper triangle, suggesting that the order effects did not change qualitatively the similarity ratings.

Expertise analysis. Testing if musically trained participants emphasised structural information more than untrained listeners began with calculating the zero-order correlations between each participant's ratings and the theoretical predictors. This gave 34 coefficients for each variable (contour, rhythm, tonality, and metre), indicating how influential each variable was for each participant. The second step was to correlate these values with years of musical training, which revealed how the contribution of each variable changed as a function of expertise. The strongest of these correlations was a trend toward greater sensitivity to contour for musically trained participants, $r(32) = .31$, $p = .077$ (two-tailed), but not rhythm, $r(32) = .01$, $p = .958$. There was no significant association between expertise and use of tonality, $r(32) = -.26$, $p = .130$, nor between expertise and metre, $r(32) = -.20$, $p = .267$. In other words, musically trained participants trended towards better use of surface information in their ratings of perceived melodic similarity, but not for any other variable (time surface, pitch structure, time structure).

Categorical ANOVA analysis. Testing for categorical effects of surface and structure of pitch and time on similarity ratings used a 2 x 2 x 2 x 2 univariate ANOVA of contour, tonality, rhythm, and metre (for all variables, the levels were same, or different). In this analysis there were main effects of contour, $F(1, 240) = 83.7$, $p < .001$, $\eta^2 = .14$ and rhythm, $F(1, 240) = 190.3$, $p < .001$, $\eta^2 = .31$, but not tonality, $F(1, 240) = 2.2$, $p = .134$, $\eta^2 < .01$, nor metre, $F(1, 240) = 1.2$, $p = .273$, $\eta^2 < .01$[4]. That is, both surface variables were significant, but neither structural variable was. Only one interaction met the threshold of significance – between rhythm and metre, $F(1, 240) = 3.9$, $p = .050$, $\eta^2 < .01$, reflecting the pattern that preserving metre only raised perceived similarity when the rhythm was different between melodies (metre had no effect when rhythm stayed constant) – see Table 3. Contour and tonality approached a significant interaction, $F(1, 240) = 3.4$, $p = .066$, $\eta^2 < .01$, and followed the opposite pattern of surface and structure, that is, tonality marginally increased similarity only when the contour was the same, and was completely ineffective when contour changed. Table 4 shows this (non-significant) pattern. No other interactions were significant (all *F*s < 1).

Figure 3 shows the similarity ratings for all pitch level combinations averaged across participant and across time levels (i.e., all combinations of the four pitch manipulation levels); Figure 4 provides the complement for time. These figures show the same relative patterns of perceived similarity, such that the values along the ascending diagonal (matching levels) are most similar, with decreasing similarity toward the opposite corners.

A potential concern from Figure 4 is that participants may have been unable to differentiate between the first two levels of temporal manipulation (t1: original; t2: ametric original rhythm). Indeed, when presented with irregular timing intervals, listeners tend to regularise them to a standard metrical framework (Motz, Erickson, & Hetrick, 2013; Repp, London, & Keller, 2011). Figure 2 shows high similarity between p1t1-p1t2 (and the reverse order), and Figure 4 shows high similarity between t1 and t2 variants. However, the confidence intervals associated with Figure 4 show that participants gave significantly higher similarity ratings to t1-t1 pairs (*M* = 5.50, CI [5.32, 5.69]) than t1-t2 (*M* = 5.19, CI [4.99 5.39]) and t2-t1 pairs (*M* = 5.12, CI [4.93 5.30]); see also factor analysis scores described below and depicted in Figure 5.

Following the main effect of order observed in the overall data, the difference between mean similarity ratings of the lower triangle (4.1, *SD* = .65) and upper triangle (4.3, *SD* = .58) of Figure 3 (pitch variant levels), was significant, $t(33) = 3.3$, $p = .002$. Also as before, the pattern of similarity across levels was alike: the upper and lower triangles of Figure 3 correlate at $r(4) = .79$. For time, the similarity ratings of the cells in the lower triangle (*M* = 4.0, *SD* = .69) of Figure 4 were also lower than those in the upper triangle (*M* = 4.4, *SD* = .66), $t(33) = 3.4$, $p = .002$. The pattern of ratings correlated highly across order, $r(4) = .88$. These analyses reaffirm that although the range of similarity ratings varied across order (i.e., there was a main effect of order), the pattern across pitch and time levels remained the same.

Regression analysis. The second analysis approach of linear regression equation predicted the similarity ratings using measures of contour, rhythm, tonality, and metre (see Data analyses section of Method for details). As signed predictors of tonality and metre were not related to ratings (*r* = -.05 and -.06, respectively), only the absolute difference values were included in the regression. Table 5 shows the final equation, which explained 60% of the variance using the rhythm, contour, and metricality predictors (in order of contribution strength), with the expected coefficient sign. Tonality was not a significant predictor of perceived similarity, despite a significant zero-order correlation *r* = -.18, *p* = .005. The number of shared pitch classes between melodies also did not explain any of the variance in ratings (by design). Four multiplicative interaction predictors were also tested: contour and rhythm (pitch and time surface), tonality and metre (pitch and time structure), contour and tonality (pitch surface and structure), as well as rhythm and metre (time surface and structure). None contributed any unique variance beyond the existing predictors.

Factor analysis. Four factors (all eigenvalues above 1) explained 87% of the variance in the ratings (see Table 6 for factor scores). To interpret the identity of factors, the factor scores were compared with the predictors from the regression equation (contour, tonality, rhythm, and metre), following Eerola et al. (2001). This required converting the factor scores into distances by calculating pairwise differences between all possible combinations of the 16 variants (p1t1-p1t2, p1t1-p1t3,..., p4t4-p4t4), yielding a 256-element vector for each factor. The absolute value of these distances (higher numbers representing greater distance in the factor space) was then correlated with the regression predictors, giving the values shown in Table 7. The highest correlations (in bold) denote the predictor with which each factor correlated best[5], which turned out to be rhythm, contour, metre, and tonality, respectively.

Figures 5 and 6 provide a visualisation of the four factor scores as 2-dimensional similarity maps. The coordinates from the temporal factors of rhythm and metre (factors 1 and 3) are depicted in Figure 5; Figure 6 shows the coordinates from the pitch factors of contour and tonality (factors 2 and 4). Overall, the factor analysis demonstrates that the four independent extracted factors correspond to the surface and structural stimulus manipulations, and explain the perceived similarity ratings remarkably well.

Discussion

The factorial manipulations of melodic surface and structure in both pitch and time (contour, rhythm, tonality, and metre) allowed investigation of their respective and combined roles in perceived similarity. There were four results of particular importance. First, although both surface and structural information contributed to ratings, rhythm and contour (surface information) were the primary determinants of perceived similarity. Second, order effects were slight and theoretically inconsequential. Third, despite the central role of pitch in Western music, temporal factors were the stronger predictors of melodic similarity in all analyses. Fourth, the predictors functioned essentially independently.

The predominance of surface information is consistent with findings in the perception of unfamiliar music (Eerola et al., 2001; Halpern, 1984; Lamont & Dibben, 2001; McAdams et al., 2004). It is likely that with increased exposure to the same melodies, structural information would become a stronger contributor to perceived similarity, as previous authors have demonstrated (Pollard-Gott, 1983; Serafine et al., 1989). However it seems unlikely that increased musical training would play a role, as expertise was not associated with greater sensitivity to either form of structure. If anything, musicians were slightly better at noticing contour (surface) changes, but not at the expense of structural information. This finding is more consistent with a generalised increase in ability to process melodic information due to greater skill in musical tasks.

Asymmetries in similarity ratings can occur when one stimulus is less structured than the other (Garner, 1974). Bartlett and Dowling (1988) observed this effect in a musical context when comparing tonal (structured) and atonal (unstructured) melodies. The current experiment shows a consistent pattern in that similarity ratings were lower when the less-altered variant occurred first, but it seems not to have affected the contributions of surface and structure in pitch and time, as the patterns were alike on either side of the diagonals of Figures 2-4. Thus there was an overall magnitude change in ratings based on order, which did not alter *how* listeners evaluated similarity in theoretical terms. That is, order effects occurred, but there is no evidence that they influenced the main theoretical question of interest, which is how listeners used contour, tonality, rhythm, and metre in rating melodic similarity.

The extent of the predominance of temporal variables is striking, given the fundamental importance of contour and tonality in music perception (Dowling, 1978; Schmuckler, 2004). As discussed in the Introduction, most work has found that temporal features dominate melodic similarity ratings of unfamiliar melodies, but there are reports of pitch being more important in similarity ratings and recognition (e.g., Carterette, Kohl, & Pitt, 1986; Hébert & Peretz, 1997; Jones et al., 1987).

Pitch and time were independent in this experiment. Of the 11 interaction terms in the ANOVA (six 2-way, four 3-way, one 4-way), only the two within-dimension terms even approached significance (rhythm-metre and contour-tonality). Additionally, none of the regression interaction terms were significant. By itself, the existence of four factors in the factor analysis does not provide evidence of independence because the technique is specifically designed to extract independent predictors. Nonetheless, the fact that 87% of the total variance was explained with these independent factors that mapped well onto the manipulations reinforces the independence found in the other analyses. These mappings were not perfect, as occasional points were counterintuitive (p1t2 is on the wrong side of the x axis in Figure 5, as are p1t3 and p3t1 in Figure 6). These exceptions represent conflicts with the accordingly weaker (i.e., structural) dimensions.

Variations in observed independence or interaction of pitch and time may stem from unequal discriminability (Garner & Felfoldy, 1970), or one dimension being more salient than another (Prince, Thompson, et al., 2009). Indeed, sufficiently imbalanced dimensional salience (e.g., via changes in stimulus structure, or task) can obscure otherwise observable pitch-time interactions (Prince, 2011; Prince, Schmuckler, & Thompson, 2009). Perhaps in this experiment temporal variables were sufficiently stronger than pitch variables so as to suppress any observable interaction between dimensions. In particular, the fact that the melodies all had the same tempo means that both relative and absolute timing information was available for use in evaluating similarity. For example, a p3t2-p4t2 comparison had not only the same sequence of duration ratios, but *exactly* the same durations themselves. Using a constant tempo was a deliberate choice in Experiment 1, so that the total elapsed duration of both melodies in a pair remained constant. In comparison, transposing the melodies to different keys preserved only the relative pitch patterns, not the exact pitch classes. Therefore the temporal dimension was in a sense more reliable, providing more stable cognitive reference points for listeners to use in rating melodic similarity.

Transposition provided a further handicap to the pitch dimension of these melodies, because after hearing the first melody in one key, listeners then had to reorient to a new key when the second melody started. Even if both melodies in the pair are tonal, the second melody will sound atonal until the listener adjusts to the new key, decreasing the perceived similarity accordingly; in most cases there would also be carryover effects onto the first melody of the next trial. Thus the effects of transposition may have decreased the informative value of pitch, causing a relative increase in salience of time. In turn, a sufficiently high imbalance in dimensional salience may have reduced the chance of observing pitch-time interactions.

Experiment 2 tested the effects of tempo change and transposition on perceived melodic similarity in order to address these issues and further explore how listeners use pitch and time in this context.

Experiment 2

There were two alterations to the Experiment 1 stimulus melodies in Experiment 2. First, the two melodies within a trial were played at different speeds. Second, a chord cadence preceded each melody, which established both the upcoming key and tempo before the melody itself began. One result of these changes is that listeners had only relative timing and relative pitch cues to evaluate similarity, instead of also preserving absolute timing information. Another important implication is that by establishing both the new key and tempo before the melody started, structure-preserving variants would not appear as unstructured, having adjusted to the new tonal centre and metrical framework before the melody started. There were no other changes to the stimuli or the experimental design.

Method

Participants. A new set of 34 participants were recruited for Experiment 2, with an average age of 25.9 (*SD* = 8.5), and 2.8 years of musical training (*SD* = 4.1). Participants were again recruited from the Murdoch University community, and provided with modest financial compensation or course credit.

Stimuli. As noted above, stimuli were the same melodies from Experiment 1, but with a chord cadence prefix and at one of two different tempi. The faster melodies were from Experiment 1; slower versions were added (2/3 the speed of the fast melodies). The durations of each note in the slower melodies were either 250 ms (eighth note), 500 ms (quarter note), 750 ms (dotted eighth note), or 1000 ms (half note). The chord cadence (I-V-I cadence, transposed to the appropriate key) was always tonal, to prevent a confound with the tonality manipulation of the melody (see Figure 7). Tempo order (slow-fast or fast-slow) was counterbalanced throughout the experiment and across participant. Minor mode melodies had a harmonic minor I-V-I cadence.

Procedure. The Experiment 2 procedure was the same as that of Experiment 1. Participants were instructed to rate the similarity of the melodies and disregard the chord cadence prefix. They waited an average of 1.0 s between melodies (*SD* = .36, median = .97). Due to the longer stimuli and slower melodies, average completion time increased to 45 minutes.

Data analysis. The data analysis approaches were the same as in Experiment 1.

Results

Preliminary checks. The average rating for the 16 match conditions was 5.04 (*SD* = .64), compared to 3.97 (*SD* = .75) for the 240 non-match conditions, indicating that participants were able to complete the task successfully, $t(33) = 10.1$, $p < .001$. Figure 8 is a greyscale plot of the 16 x 16 similarity matrix, showing as in Experiment 1 that the diagonal (match) conditions received the highest similarity ratings, which decreased away from the diagonal. The axes have been reordered from Experiment 1 in accordance with which dimension was more influential (time for Experiment 1, pitch for Experiment 2). This reordering does not change any analyses, but is intended to display the decreasing similarity from the diagonal (match conditions) more clearly.

Order effects. The ratings of the lower triangle (more stable variant first) conditions (*M* = 3.8, *SD* = .79) were significantly lower than the upper triangle conditions (*M* = 4.2, *SD* = .76), $t(33) = 5.0$, $p < .001$, indicating the presence of an order effect. The correlation between upper and lower triangles ($r = .27$) was lower than the subgroup intercorrelations ($r = .44$; .56), when calculated as described in Experiment 1.

Expertise analysis. The same analysis technique as in Experiment 1 revealed that musical training enhanced the use of contour, $r(32) = .39$, $p = .021$ (two-tailed), but not rhythm, $r(32) = -.22$, $p = .202$. There was no significant association between expertise and use of tonality, $r(32) = -.07$, $p = .679$, nor between expertise and metre, $r(32) = -.251$, $p = .152$. Thus again musically trained participants were better able to make use of contour information in perceived similarity ratings, but no other differences emerged across expertise.

Categorical ANOVA analysis. The 2 x 2 x 2 x 2 univariate ANOVA testing the effects of pitch and time manipulations on similarity ratings revealed main effects of contour, $F(1, 240) = 60.2$, $p < .001$, $\eta^2 = .12$, rhythm, $F(1, 240) = 47.1$, $p < .001$, $\eta^2 = .09$, and tonality, $F(1, 240) = 11.0$, $p = .001$, $\eta^2 = .02$, but not metre, $F(1, 240) = 1.0$, $p = .309$, $\eta^2 < .01$. Only the interaction between contour and tonality was significant, $F(1, 240) = 4.0$, $p = .046$, $\eta^2 = .01$; the rhythm-metre interaction approached but did not reach the threshold, $F(1, 240) = 3.2$, $p = .075$, $\eta^2 = .01$. No other interactions were significant. The similarity ratings associated with the main effect of pitch manipulations (averaged across time levels) are depicted in Figure 9; Figure 10 has the same for time.

Listeners differentiated more between pitch variants when the less-altered variant was heard first – that is, the lower triangle of Figure 9 has overall lower similarity ratings ($M = 4.1$, $SD = .65$) than the upper triangle ($M = 4.3$, $SD = .58$), $t(33) = 4.5$, $p < .001$. Similarly for time, the lower triangle of Figure 10 received significantly lower ratings than the upper triangle (M lower = 3.7, $SD = .73$; M upper = 4.2, $SD = .87$), $t(33) = 5.3$, $p < .001$. For both pitch and time, the upper and lower triangles were positively correlated, $r(4) = .34$ and .35, respectively, showing agreement across order (albeit less than in Experiment 1).

Regression analysis. Regressing objective similarity predictors on the 256 similarity ratings averaged across participant gave significant effects of contour, rhythm, and tonality (absolute, not signed), but not metre; see Table 8 for the equation details. Additionally, the number of shared pitch classes between melody pairs was not a significant predictor of similarity ratings. In total, the equation accounted for 35% of the variance, less than the 60%

of Experiment 1. No interaction terms accounted for additional variance. Note that pitch factors now had a much stronger relative contribution to ratings than the previous experiment.

Factor analysis. Principal components factor analysis of the 16 x 16 similarity matrix yielded 5 factors with eigenvalues greater than 1 (see Table 9 for factor scores). Only the first three factors were interpretable, as contour, tonality, and rhythm, in order of variance accounted for (see Table 10 for correlations between factor scores and regression predictors). Together these factors account for 61% of the variance (using the full five-dimensional solution accounts for 77%). Figure 11 graphs the variants in a 2-dimensional similarity space based on pitch (factors 1 and 2), whereas Figure 12 depicts rhythm (factor 3) along with the unexplained factor 4.

Discussion

Experiment 2 investigated the roles of surface and structural information in pitch and time on similarity ratings when sequences in a given trial were transposed in tempo as well as key, both of which were prepared by chord cadence prefixes. Thus only relative information (not absolute) was available in both pitch and time, providing a more conceptually equal basis for comparison between dimensions. This alteration to the task made it longer and harder, resulting in more noise in the data; additionally, order effects were stronger than in Experiment 1. Nevertheless, three findings of interest emerged. First, the balance of predictive value shifted away from temporal variables (rhythm and metre) and toward pitch information (contour and tonality). Second, surface information remained more prominent than structure (and expertise did not modify this balance). Third, despite more equalised effects of pitch and time, no interactions between the dimensions emerged.

Compared to Experiment 1, the data of Experiment 2 exhibited larger standard deviations, lower intersubject correlations, smaller ANOVA effect sizes, and smaller amounts of variance explained in regression and factor analysis techniques. Together, these findings all point toward Experiment 2 being more difficult than Experiment 1 for participants. Accordingly, removing the absolute timing information by introducing a tempo change substantially reduced participants' ability to judge the similarity of the variants, which follows from findings that tempo changes impair memory for melodies (Dowling, Bartlett, Halpern, & Andrews, 2008; Halpern & Mullensiefen, 2008; Schellenberg et al., 2013).

Order effects were more observable in Experiment 2, as evidenced by lower similarity ratings for conditions that presented a less-altered melody first (i.e., the lower vs. upper triangles of Figures 8-10). The order effects show that participants seem to have had greater trouble differentiating between sequences when the first sequence was a variant that departed more from the original (as in Bartlett & Dowling, 1988; Watkins, 1985). This pattern is especially notable in Figure 10, which depicts the main effects of comparing each time level manipulation. In this figure, neither levels t3 nor t4 receive the highest similarity on their diagonal (matching) conditions – that is, the t3-t3 conditions were rated as less similar than t3-t1 conditions; likewise, t4-t1 and t4-t2 conditions receive higher similarity ratings than t4-t4. In all other corresponding figures (3, 4, and 9), the match conditions were always the most similar.

In Experiment 2, listeners relied more on pitch than time to form their similarity judgements. Specifically, contour was the most influential predictor whereas rhythm was weaker; tonality was stronger than previously observed while metre no longer contributed to ratings at all. The stronger role of tonality compared to Experiment 1 may represent the removal of the penalty on the dimension of pitch from an unprepared transposition, as transposing a melody to a new key makes recognition more difficult (Dowling & Fujitani, 1971; Stalinski & Schellenberg, 2010).

In both experiments, surface information was generally a more powerful predictor of perceived similarity than structural information, although tonality (pitch structure), rivalled that of rhythm (time surface) in Experiment 2. Thus simply inserting a chord cadence before each melody resulted in a stronger role of structure in Experiment 2, even without repeated presentations of these unfamiliar melodies, which is known to affect the balance between surface and structure (Pollard-Gott, 1983; Serafine et al., 1989). Again musical expertise aided the participants in recovering contour information when forming their rating, but it did not alter the relative emphasis on surface and structure.

The main goal of Experiment 2 was to see if putting pitch and time on a more conceptually equal footing altered their relative contribution to perceived similarity, and if interactions between pitch and time were more apparent when the effect sizes were more equal. From the ANOVA effect sizes and the squared semipartial correlations of the regression, the experiment was successful in altering the weighting of the dimensions toward parity, as pitch and time accounted for a more comparable amount of variance in the data. Nevertheless, pitch-time interactions were no stronger than in the previous experiment, and the factor analysis again retrieved independent factors that corresponded well to the manipulations. This finding aligns with recent findings of independence of tonal information and tempo (Schellenberg et al., 2013). In terms of dimensional salience, these data argue for independence as no interaction emerged despite equalisation of (or at least changes to) relative main effect sizes (Prince, 2011).

General Discussion

Two experiments explored how systematic manipulations of pitch and temporal surface and structural information (contour, rhythm, tonality, and metre) affected similarity ratings of unfamiliar melodies. In both experiments, surface information was the more effective predictor of ratings. Temporal factors were more influential in Experiment 1, possibly because tempo was constant, enabling the use of both absolute and relative timing cues in similarity judgements, whereas only relative pitch information was available due to the use of transposition. In Experiment 2, the tempo of the melodies varied within-trial, transpositions were prepared with chord cadence prefixes, and pitch became the dominant dimension. Neither experiment obtained an interaction between dimensions, and within-dimension interactions between surface and structure were small and unreliable.

This detailed examination of the role of surface and structure in both pitch and time on melodic similarity offers several theoretical advances. First, the relative strength of pitch and temporal factors is flexible, altered dramatically by changing the tempo and adding a chord prefix. Second, the factorial design revealed independent roles of pitch and time, regardless of the relative strength of these dimensions. Third, interactions between surface and structural information were unreliable and never occurred between dimensions. Subsidiary findings came from the minimal role of expertise and overall effects of melody order.

In Experiment 1, the influence of rhythm and metre on similarity ratings dominated that of contour and tonality. Adding a chord cadence and varying the tempo between melodies (Experiment 2) reversed this bias. This finding that judgements of melodic similarity are flexible is entirely new, as far as the author is aware. One explanation is that listeners rely more on absolute timing information than predicted by the more conventional approach that listeners encode temporal patterns in relative terms instead of absolute (e.g., Drake & Botte, 1993; Miller & McAuley, 2005). This is not the only evidence of memory for absolute duration in musical sequences, as Levitin and Cook (1996) found that 72% of untrained listeners’ vocal productions of popular melodies were within 8% of the actual tempo (ruling out effects of articulatory constraints). Moreover, there is evidence of the importance of absolute information in melodic memory and recognition, such as tempo (Halpern, 1988; Halpern & Mullensiefen, 2008; Schellenberg et al., 2013), timbre (Lange &

Czernochowski, 2013; Poulin-Charronnat et al., 2004), key or pitch height (Creel, 2011; Halpern, 1989; Schellenberg & Trehub, 2003), and articulation (Wee Hun Lim & Goh, 2013).

However, the importance of absolute timing information does not explain the change in weighting of contour, rhythm, tonality, and metre – this information was the same in both experiments, so why did their relative importance vary? A possible explanation derives from dimensional salience (Prince, Thompson, et al., 2009), where the informative value of a dimension can prioritise it at the expense of other dimensions, even when entirely irrelevant to the task. This prioritisation affects the dimension as a whole, such that in this case, all temporal information – including rhythm and metre – became more important when the dimension was more reliable by virtue of preserving the absolute timing relations. Previous work on varying the salience of a dimension has focused on pitch manipulations (Prince, 2014; Prince, Schmuckler, et al., 2009); the current findings suggest that the preservation of tempo can have a similar effect on the salience of time.

The factorial design of the experiments (as recommended by Eerola et al., 2001), enabled stronger evaluations of the relations between manipulated variables. Additionally, the large set size and counterbalancing arrangements vastly reduced the potential role of familiarity with the melodies. In this context, the findings were consistent with complete independence between pitch and time, as there were no significant interactions between these dimensions in either experiment.

The litany of conflicting reports in the literature on pitch-time integration demonstrates that they cannot be purely independent or interactive (for reviews, see Prince, 2011; Schellenberg et al., 2013). Instead, the question of interest is what influences the pattern observed in a given circumstance. One potential factor is whether the nature of the task fosters a global or local style of processing, whereby global processing makes interactive relations more likely (Jones & Boltz, 1989; Tillmann & Lebrun-Guillaud, 2006). Melodic similarity is a global task, and yet there were independent contributions of both pitch and time in these data. Preserving the original first and last pitches/durations in all variants was specifically intended to reduce the effectiveness of a local strategy such as focussing on one section of the melody (e.g., the beginning or end). The stage model of pitch-time interactions proposes that the dimensions are initially separate, and recombined at a later stage in processing (Peretz & Zatorre, 2005; Thompson, Hall, & Pressing, 2001). The present data do not fit this approach either – evaluating melodic similarity clearly involves late stages of processing and yet independence remained.

Event coherence may influence how participants attend to sequences (Jones & Boltz, 1989). Events are coherent if the pitch accent structure coincides with the temporal accent structure, for instance ending a melodic phrase (pitch accent) with a lengthened note (temporal accent); this leads to greater integration of pitch and temporal information than if they are offset. Melodies with coherent pitch-time accent structures are recognised (same-different) more accurately than if they are incoherent (Boltz, 1998; Schulkind, 1999). In the present experiments, the coincidence of pitch and temporal accent structures was not systematically controlled, although it almost certainly varied. Thus perhaps listeners did not adopt an integrated attending style, as only 1/16 of the melodies (the p1t1 variants) would have had a consistently coherent accent structure. Still, if integrating pitch and time is subject to a number of conditions (coherence, task difficulty, task design, etc.), then perhaps it is afforded only by particular contexts rather than the default listening mode.

Attenuating the difference between the main effect sizes of pitch and temporal factors did not uncover an interaction between them (in Experiment 2), suggesting that any potential difference in salience between dimensions (in Experiment 1) was not obscuring an interaction between pitch and time. Research on goodness ratings of single melodies with similar manipulations to the present experiments found that pitch and time interacted when the main effect sizes (squared semipartial correlations) of the two dimensions were more equal (Prince, 2011). That finding was consistent with an interpretation that interactions between dimensions can become obscured when one dimension is more salient than another (Prince, Schmuckler, et al., 2009). Yet in the current experiments, there was no hint of an interaction, even when the effect sizes were nearly identical.

Testing the role of surface and structure in melodic similarity was among the primary foci of the present study. The overall superiority of surface information over structure in similarity ratings of unfamiliar melodies aligns with previous research, but examining exactly how these forms of information combine is unique to this research. Although ANOVA analyses suggested subtle interactions between surface and structure, these occurred only within a given dimension (contour with tonality, and rhythm with metre), and were unreliable and inconsistent (see Tables 3 and 4). No interactions involved both dimensions (such as contour with rhythm), giving further weight to the observed independence of pitch and time. The rhythm-metre interaction was unreliable (only in Experiment 1), and counterintuitive – suggesting that metre was more influential on ratings when the two melodies had a different rhythm. This pattern was inconsistent with the contour-tonality interaction, which was that tonality only contributed to perceived similarity when the contour remained the same in a given melody pair. Although this interaction was more consistent across experiment, it emerged only in the ANOVA analysis, just barely reached significance, and only in Experiment 2. Overall, surface and structure were largely independent in these experiments, and particularly so across dimension, which is a novel finding in melodic similarity research.

The present sample was representative of an undergraduate population in that the average experience was around 3 years of musical training (mode = 0). Expertise had little bearing on the findings, in that greater musical training afforded improved ability to use contour – but not at the expense of other information. It is possible that highly trained musicians would employ a more analytic strategy and demonstrate greater sensitivity to structural information (Bigand, 1997; Krumhansl & Shepard, 1979). For instance, Frankland and Cohen (1996) found that a tonal context improved pitch height comparison accuracy for musicians only, but there is also evidence of both populations benefitting from tonality (Schulze, Dowling, & Tillmann, 2012). In the context of melodic recognition, musicians can also be more flexible in their reliance on tonal or contour information, but moderately experienced listeners also use both types (Dowling, 1986). Musicians use a slightly different brain network for remembering tonal sequences, consistent with a more exact representation of the sequence (Schulze, Mueller, & Koelsch, 2011). But on the whole, all listeners show remarkably similar perceptual processes and neural activity in response to music (Bigand & Poulin-Charronnat, 2006; Koelsch, Gunter, Friederici, & Schröger, 2000), including studies of perceived similarity (Halpern, 1984; Halpern et al., 1998). Moreover, expertise appears not to change the pattern of pitch-time integration (Boltz, 1989; Hébert & Peretz, 1997; Lebrun-Guillaud & Tillmann, 2007; Palmer & Krumhansl, 1987; Smith & Cuddy, 1989; Tillmann & Lebrun-Guillaud, 2006). If anything, musicians may be better at separating the dimensions and attending selectively (Pitt & Monahan, 1987), but given that the current findings already showed independent effects, it is unlikely that increased musical expertise would change the observed pitch-time integration pattern. Addressing this question specifically will require direct comparisons between highly trained and untrained listeners, similar to more general research on the use of surface and structure in physics experts and novices (Chi et al., 1981).

Another potential influence on the results comes from order effects – that is, the same stimuli sometimes received different ratings depending on the order of their presentation. Similarity ratings were lower when the first sequence of a trial adhered more to the original melody's characteristics (e.g., p1t1-p4t1 vs. p4t1-p1t1 in Experiment 2 – see Figure 8), in agreement with previous findings in both pitch (Bartlett & Dowling, 1988; Watkins, 1985) and time (Bharucha & Pryor, 1986; Kidd, Boltz, & Jones, 1984). Bartlett and Dowling showed that their order effects were not due to greater memorability of more tonal melodies, and Schellenberg (2002) explained that “going out-of-tune is more noticeable than going in-tune” when discriminating musical intervals. Both follow from Krumhansl’s (1979, 1990) formalisation of contextual asymmetry – pitches not belonging to the current musical key are judged as more similar to pitches inside the key than vice versa. For both experiments in the present study, order effects were larger for time manipulation levels (Figures 4 and 10) than pitch manipulation levels (Figures 3 and 9). However the same basic patterns emerge on either side of the diagonal – similarity ratings generally decrease with increasing distance from the diagonal. Overall, the order effects seem to indicate that the task was easier when the less-violated sequence occurred first, but there were no observable qualitative differences in how listeners used and combined the surface and structural information across order.

There are limitations to this research that warrant contemplation. First, all of the seed melodies used only the most common metrical framework of Western music (4/4 time). It is unlikely that the roles of pitch and temporal surface and structure function in a radically different way for other metres (cf. Smith & Cuddy, 1989), but is untested in a melodic similarity context.

Second, the dichotomous manipulations of tonality and metre (tonal/atonal and metric/ametric) do not capture more nuanced manipulations of tonality and metre, such as contrasting the same pitch sequence in different metrical frameworks. Likewise, the major and minor forms of tonality are perceptually subtle (Halpern, 1984; Halpern et al., 1998) but structurally significant.

Third, the manipulations in the stimuli may not have been as independent as hoped, particularly rhythm and metre. Rearranging the durations of a sequence can affect its metrical strength (Povel & Essens, 1985); in this case the randomly reordered t3 variants had lower metricality than t1, as measured by correlation with the Palmer and Krumhansl (1990) metric hierarchy. However, the present regression analysis used separate and continuous predictors of metre and rhythm, and there were no significant interaction terms; moreover the factor analysis supported independent roles for these predictors (see Figure 5). Regardless, future work may be able to separate rhythmic and metrical information more effectively.

In closing, the perceived similarity between two novel melodies depends on surface and structural features of pitch and time. For immediate similarity ratings of novel sequences that vary in these properties, surface information is more influential. Within a dimension, surface and structure may show small interactions, but not across dimensions. The relative contribution of pitch and time is flexible based on the availability of relative versus absolute information, but these dimensions function independently regardless of which is stronger.

References

Abe, J.I., & Okada, A. (2004). Integration of metrical and tonal organization in melody perception. *Japanese Psychological Research, 46*, 298-307.

Acevedo, S., Temperley, D., & Pfordresher, P.Q. (2014). Effects of metrical encoding on melody recognition. *Music Perception: An Interdisciplinary Journal, 31,* 372-386.

Barnes, R., & Johnston, H. (2010). The role of timing deviations and target position uncertainty on temporal attending in a serial auditory pitch discrimination task. *Quarterly Journal of Experimental Psychology, 63,* 341-355. doi: 10.1080/17470210902925312

Barnes, R., & Jones, M.R. (2000). Expectancy, attention, and time. *Cognitive Psychology, 41*, 254-311.

Bartlett, J.C., & Dowling, W.J. (1988). Scale structure and similarity of melodies. *Music Perception, 5,* 285-314.

Bharucha, J.J., & Pryor, J.H. (1986). Disrupting the isochrony underlying rhythm: An asymmetry in discrimination. *Perception & Psychophysics, 40*, 137-141. doi: 10.3758/BF03203008

Bigand, E. (1997). Perceiving musical stability: The effect of tonal structure, rhythm, and musical expertise. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 808-822.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition, 100*, 100-130.

Boltz, M.G. (1989). Perceiving the end: Effects of tonal relationships on melodic completion. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 749-761.

Boltz, M.G. (1998). The processing of temporal and nontemporal information in the remembering of event durations and musical structure. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1087-1104.

Boltz, M.G. (2011). Illusory tempo changes due to musical characteristics. *Music Perception, 28*, 367-386.

Bregman, A.S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA, US: MIT Press.

Carterette, E.C., Kohl, D.V., & Pitt, M.A. (1986). Similarities among transformed melodies - the abstraction of invariants. *Music Perception, 3*, 393-410.

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152. doi: 10.1207/s15516709cog0502_2

Creel, S.C. (2011). Specific previous experience affects perception of harmony and meter. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1512-1526. doi: 10.1037/a0023507

Cuddy, L.L., & Cohen, A.J. (1976). Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology, 28*, 255-270.

Cuddy, L.L., Cohen, A.J., & Mewhort, D.J. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 869-883.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception, 32*, 341-365. doi: 10.1068/p3370

Deutsch, D. (Ed.). (2013). *The psychology of music* (3rd ed.): Academic Press.

Dewitt, L.A., & Crowder, R.G. (1986). Recognition of novel melodies after brief delays. *Music Perception, 3*, 259-274.

Dowling, W.J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics, 14*, 37-40.

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

Dowling, W.J. (1986). Context effects on melody recognition - scale-step versus interval representations. *Music Perception, 3*, 281-296.

Dowling, W.J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics, 50*, 305-313.

Dowling, W.J., Bartlett, J.C., Halpern, A.R., & Andrews, M.W. (2008). Melody recognition at fast and slow tempos: Effects of age, experience, and familiarity. *Perception & Psychophysics, 70*, 496-502.

Dowling, W.J., & Fujitani, D.S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49*, 524-531.

Dowling, W.J., Kwak, S., & Andrews, M.W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics, 57*, 136-149.

Drake, C., & Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Attention, Perception, & Psychophysics, 54*, 277-286. doi: 10.3758/bf03205262

Edworthy, J. (1985). Interval and contour in melody processing. *Music Perception, 2*, 375-388.

Eerola, T., Järvinen, T., Louhivuori, J., & Toiviainen, P. (2001). Statistical features and perceived similarity of folk melodies. *Music Perception, 18*, 275-296. doi: 10.1525/mp.2001.18.3.275

Eiting, M.H. (1984). Perceptual similarities between musical motifs. *Music Perception, 2*, 78-94.

Essens, P.J. (1986). Hierarchical organization of temporal patterns. *Perception & Psychophysics, 40*, 69-73. doi: 10.3758/bf03208185

Essens, P.J., & Povel, D.J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics, 37*, 1-7.

Freedman, E.G. (1999). The role of diatonicism in the abstraction and representation of contour and interval information. *Music Perception, 16*, 365-387.

Garner, W.R. (1970). Good patterns have few alternatives. *American Scientist, 58*, 34-&.

Garner, W.R. (1974). *The processing of information and structure* (Vol. 203). Oxford, England: Lawrence Erlbaum.

Garner, W.R., & Felfoldy, G.L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1*, 225-241.

Garner, W.R., & Gottwald, R.L. (1968). Perception and learning of temporal patterns. *Quarterly Journal of Experimental Psychology, 20*, 97-109.

Halpern, A.R. (1984). Perception of structure in novel music. *Memory & Cognition, 12*, 163-170.

Halpern, A.R. (1988). Perceived and imagined tempos of familiar songs. *Music Perception, 6*, 193-202.

Halpern, A.R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition, 17*, 572-581.

Halpern, A.R., Bartlett, J.C., & Dowling, W.J. (1998). Perception of mode, rhythm and contour in unfamiliar melodies: Effects of age and experience. *Music Perception, 15*, 335-355.

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

Handel, S., & Oshinsky, J.S. (1981). The meter of syncopated auditory polyrhythms. *Perception & Psychophysics, 30*, 1-9. doi: 10.3758/bf03206130

Hébert, S., & Peretz, I. (1997). Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? *Memory & Cognition, 25*, 518-533.

Jones, M.R., & Boltz, M.G. (1989). Dynamic attending and responses to time. *Psychological Review, 96*, 459-491.

Jones, M.R., & Ralston, J.T. (1991). Some influences of accent structure on melody recognition. *Memory & Cognition, 19,* 8-20.

Jones, M.R., Summerell, L., & Marshburn, E. (1987). Recognizing melodies - a dynamic interpretation. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 39*, 89-121.

Kidd, G., Boltz, M.G., & Jones, M.R. (1984). Some effects of rhythmic context on melody recognition. *American Journal Of Psychology, 97,* 153-173.

Koelsch, S., Gunter, T.C., Friederici, A.D., & Schröger, E. (2000). Brain indices of music processing: "Nonmusicians" are musical. *Journal of cognitive neuroscience, 12*, 520-541.

Krumhansl, C.L. (1979). Psychological representation of musical pitch in a tonal context. *Cognitive Psychology, 11*, 346-374.

Krumhansl, C.L. (1990). *Cognitive foundations of musical pitch*. New York, NY: Oxford University Press.

Krumhansl, C.L. (1991). Memory for musical surface. *Memory & Cognition, 19*, 401-411.

Krumhansl, C.L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin, 126*, 159-179.

Krumhansl, C.L., & Cuddy, L.L. (2010). A theory of tonal hierarchies in music. In M. R. Jones, R. R. Fay & A. N. Popper (Eds.), *Music perception. Springer handbook of auditory research*. New York, NY: Springer Science.

Krumhansl, C.L., & Schmuckler, M.A. (1986, July). *Key-finding in music: An algorithm based on pattern matching to tonal hierarchies.* Paper presented at the 19th Annual Meeting of the Society of Mathematical Psychology, Cambridge, MA.

Krumhansl, C.L., & Shepard, R.N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 579-594.

Kruskal, J.B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage Publications.

Lamont, A., & Dibben, N. (2001). Motivic structure and the perception of similarity. *Music Perception, 18*, 245-274.

Lange, K., & Czernochowski, D. (2013). Does this sound familiar? Effects of timbre change on episodic retrieval of novel melodies. *Acta Psychologica, 143*, 136-145. doi: 10.1016/j.actpsy.2013.03.003

Lebrun-Guillaud, G., & Tillmann, B. (2007). Influence of a tone's tonal function on temporal change detection. *Perception & Psychophysics, 69*, 1450-1459.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, Massachusetts: MIT Press.

Levitin, D.J., & Cook, P.R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics, 58*, 927-935. doi: 10.3758/bf03205494

Massaro, D.W., Kallman, H.J., & Kelly, J.L. (1980). The role of tone height, melodic contour, and tone chroma in melody recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 77-90.

McAdams, S., Vieillard, S., Houix, O., & Reynolds, R. (2004). Perception of musical similarity among contemporary thematic materials in two instrumentations. *Music Perception, 22*, 207-237.

McAuley, J.D. (2010). Tempo and rhythm. In M. R. Jones, R. R. Fay & A. N. Popper (Eds.), *Springer handbook of auditory research: Music perception* (pp. 165-199). New York, NY, US: Springer Science + Business Media.

Miller, N.S., & McAuley, J.D. (2005). Tempo sensitivity in isochronous tone sequences: The multiple-look model revisited. *Perception & Psychophysics, 67*, 1150-1160. doi: 10.3758/bf03193548

Monahan, C.B., & Carterette, E.C. (1985). Pitch and duration as determinants of musical space. *Music Perception, 3,* 1-32.

Monahan, C.B., Kendall, R.A., & Carterette, E.C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Perception & Psychophysics, 41*, 576-600.

Motz, B.A., Erickson, M.A., & Hetrick, W.P. (2013). To the beat of your own drum: Cortical regularization of non-integer ratio rhythms toward metrical patterns. *Brain and Cognition, 81*, 329-336.

Palmer, C., & Krumhansl, C.L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 116-126.

Palmer, C., & Krumhansl, C.L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 728-741.

Peretz, I., & Zatorre, R.J. (2005). Brain organization for music processing. *Annual Review of Psychology, 56*, 89-114.

Pfordresher, P.Q., & Mantell, J.T. (2014). Singing with yourself: Evidence for an inverse modeling account of poor-pitch singing. *Cognitive Psychology, 70*, 31-57. doi: 10.1016/j.cogpsych.2013.12.005

Pitt, M.A., & Monahan, C.B. (1987). The perceived similarity of auditory polyrhythms. *Perception & Psychophysics, 41*, 534-546.

Pollard-Gott, L. (1983). Emergence of thematic concepts in repeated listening to music. *Cognitive psychology, 15*, 66-94.

Poulin-Charronnat, B., Bigand, E., Lalitte, P., Madurell, F., Vieillard, S., & McAdams, S. (2004). Effects of a change in instrumentation on the recognition of musical materials. *Music Perception, 22*, 239-263.

Povel, D.J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2*, 411-440.

Prince, J.B. (2011). The integration of stimulus dimensions in the perception of music. *Quarterly Journal of Experimental Psychology, 64*, 2125-2152. doi: 10.1080/17470218.2011.573080

Prince, J.B. (2014). Pitch structure, but not selective attention, affects accent weightings in metrical grouping. *Journal of Experimental Psychology: Human Perception and Performance*. doi: 10.1037/a0037730

Prince, J.B., Schmuckler, M.A., & Thompson, W.F. (2009). The effect of task and pitch structure on pitch-time interactions in music. *Memory & Cognition, 37*, 368-381. doi: 10.3758/MC.37.3.368

Prince, J.B., Thompson, W.F., & Schmuckler, M.A. (2009). Pitch and time, tonality and meter: How do musical dimensions combine? *Journal of Experimental Psychology: Human Perception and Performance 35*, 1598-1617. doi: 10.1037/a0016456

Repp, B.H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review, 12*, 969-992.

Repp, B.H., London, J., & Keller, P.E. (2011). Perception-production relationships and phase correction in synchronization with two-interval rhythms. *Psychological Research, 75,* 227-242.

Repp, B.H., & Su, Y.H. (2013). Sensorimotor synchronization: A review of recent research (2006-2012). *Psychonomic Bulletin & Review, 20,* 403-452. doi: 10.3758/s13423-012-0371-2

Rosner, B.S., & Meyer, L.B. (1986). The perceptual roles of melodic process, contour, and form. *Music Perception, 4*, 1-40.

Schellenberg, E.G. (2002). Asymmetries in the discrimination of musical intervals: Going out-of-tune is more noticeable than going in-tune. *Music Perception, 19*, 223-248.

Schellenberg, E.G., Stalinski, S.M., & Marks, B.M. (2013). Memory for surface features of unfamiliar melodies: Independent effects of changes in pitch and tempo. *Psychological Research*, 1-12.

Schellenberg, E.G., & Trehub, S.E. (2003). Good pitch memory is widespread. *Psychological Science, 14*, 262.

Schmuckler, M.A. (1999). Testing models of melodic contour similarity. *Music Perception, 16*, 295-326.

Schmuckler, M.A. (2004). Pitch and pitch structures. In J. Neuhoff (Ed.), *Ecological psychoacoustics* (pp. 271-315). San Diego, CA: Elsevier Science.

Schmuckler, M.A. (2009). Components of melodic processing. In S. Hallam, I. Cross & M. Thaut (Eds.), *Oxford handbook of music psychology* (pp. 93-106). Oxford, UK: Oxford University Press.

Schmuckler, M.A. (2010). Melodic contour similarity using folk melodies. *Music Perception, 28*, 169-193.

Schulkind, M.D. (1999). Long-term memory for temporal structure: Evidence from the identification of well-known and novel songs. *Memory & Cognition, 27*, 896-906.

Schultz, B.G., Stevens, C.J., Keller, P.E., & Tillmann, B. (2013). The implicit learning of metrical and nonmetrical temporal patterns. *The Quarterly Journal of Experimental Psychology, 66*, 360-380. doi: 10.1080/17470218.2012.712146

Schulze, K., Dowling, W.J., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Perception, 29*, 255-267.

Schulze, K., Mueller, K., & Koelsch, S. (2011). Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *European Journal of Neuroscience, 33*, 189-196.

Serafine, M.L., Glassman, N., & Overbeeke, C. (1989). The cognitive reality of hierarchic structure in music. *Music Perception, 6*, 397-430.

Smith, K.C., & Cuddy, L.L. (1989). Effects of metric and harmonic rhythm on the detection of pitch alterations in melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 457-471.

Stalinski, S.M., & Schellenberg, E.G. (2010). Shifting perceptions: Developmental changes in judgments of melodic similarity. *Developmental Psychology*, *46*, 1799-1803.

Thompson, W.F., Hall, M.D., & Pressing, J. (2001). Illusory conjunctions of pitch and duration in unfamiliar tone sequences. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 128-140.

Tillmann, B., Gosselin, N., Bigand, E., & Peretz, I. (2012). Priming paradigm reveals harmonic structure processing in congenital amusia. *Cortex, 48*, 1073-1078.

Tillmann, B., & Lebrun-Guillaud, G. (2006). Influence of tonal and temporal expectations on chord processing and on completion judgments of chord sequences. *Psychological Research, 70*, 345-358.

Trainor, L.J., McDonald, K.L., & Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience, 14*, 430-442.

Trehub, S.E., & Hannon, E.E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition, 100*, 73-99. doi: 10.1016/j.cognition.2005.11.006

Trehub, S.E., Schellenberg, E.G., & Kamenetsky, S.B. (1999). Infants' and adults' perception of scale structure. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 965-975.

van Egmond, R., & Povel, D.-J. (1996). Perceived similarity of exact and inexact transpositions. *Acta Psychologica, 92*, 283-295. doi: 10.1016/0001-6918(95)00043-7

van Egmond, R., Povel, D.-J., & Maris, E. (1996). The influence of height and key on the perceptual similarity of transposed melodies. *Perception & Psychophysics, 58*, 1252-1259.

Watkins, A.J. (1985). Scale, key, and contour in the discrimination of tuned and mistuned approximations to melody. *Perception & Psychophysics, 37*, 275-285.

Wee Hun Lim, S., & Goh, W.D. (2013). Articulation effects in melody recognition memory. *Quarterly Journal of Experimental Psychology*, *66*, 1774-1792. doi: 10.1080/17470218.2013.766758

Author Notes

Thanks to Rachael Davey, Scott Devenish, Andrew Mannucci, Sandra O'Shea, and Taryn van Gramberg for help with Experiment 1 data collection. Please address correspondence concerning this article to Jon B. Prince, School of Psychology and Exercise Science, Murdoch University, 90 South Street, Murdoch, WA 6150 (email: j.prince@murdoch.edu.au).

Footnotes

[1] Contour and rhythm can also be considered as structural, as changing either of them may also change the identity of the melodic sequence. In the present article, these will be considered surface characteristics, if only to differentiate them from the more clearly structural variables of tonality and metre.

[2] Because each participant provided 136 ratings (not the full 16 x 16 grid), a repeated-measures ANOVA approach would have resulted in an unacceptably high number of missing cells.

[3] Using the Fourier analysis model of melodic contour (Schmuckler, 1999, 2010), two other measures of contour similarity were tested: the absolute difference score between the amplitude vectors, and the difference score between the phase vectors. Both were nearly identical to the contour correlations (r > .96 for both experiments), and thus represented the same information. For conceptual simplicity and avoidance of collinearity, the analyses use only the correlation coefficient.

[4] All η^2 values are full (not partial) eta-squared, using the corrected total type III sum of squares.

[5] The negative sign of the coefficients with rhythm and contour in Table 7 emerges because larger factor scores (i.e., greater distance) correlate negatively with these predictors, in which larger values denote greater similarity (smaller distance). Likewise, the values are positive in columns 3 and 4 as higher numbers in metre and tonality predictors indicated greater distance. In all cases the sign is consistent with the theoretical prediction. The same applies to Table 10 (Experiment 2).

Table 1

Explanation of pitch and time manipulation levels

Level	Name	Description	Surface preserved?	Structure preserved?
p1	Pitch original	Unaltered (original) sequence of pitches	Yes	Yes
p2	Atonal original contour	Pitches replaced with artificial scale (A B C# D D# F G), but retaining contour	Yes	No
p3	Contour-violated	Randomly shuffled order of original pitches, not violating tonality	No	Yes
p4	Contour-violated-atonal	Randomly shuffled order of p2 pitches	No	No
t1	Time original	Unaltered (original) sequence of durations	Yes	Yes
t2	Ametric original rhythm	Durations changed to non-metric (200, 280, 530, 650), but preserving ordinal scaling (rhythm)	Yes	No
t3	Rhythm-violated	Randomly shuffled order of original durations, unchanged metre	No	Yes
t4	Rhythm-violated-ametric	Randomly shuffled order of t2 durations	No	No

Note: the first and last pitches, as well as first and last durations, were unchanged in all conditions.

Table 2

Manipulation levels in pitch and time, and resulting condition names.

	Pitch original	Atonal original contour	Contour-violated	Contour-violated-atonal
Time original	p1t1	p2t1	p3t1	p4t1
Ametric original rhythm	p1t2	p2t2	p3t2	p4t2
Rhythm-violated	p1t3	p2t3	p3t3	p4t3
Rhythm-violated-ametric	p1t4	p2t4	p3t4	p4t4

Table 3

Interaction between rhythm and metre. Underlined values indicate conditions in which metrical similarity affected perceived similarity (see text). Standard error of the mean values are in parentheses.

	Experiment 1 Metre same	Experiment 1 Metre different	Experiment 2 Metre same	Experiment 2 Metre different
Rhythm same	5.14 (.08)	5.21 (.11)	4.38 (.08)	4.63 (.12)
Rhythm different	<u>4.19 (.08)</u>	<u>3.94 (.06)</u>	3.93 (.08)	3.86 (.07)

Table 4

Interaction between contour and tonality. Underlined values indicate conditions in which tonal similarity affected perceived similarity (see text). Standard error of the mean values are in parentheses.

	Experiment 1 Tonality same	Experiment 1 Tonality different	Experiment 2 Tonality same	Experiment 2 Tonality different
Contour same	5.12 (.08)	4.85 (.11)	<u>4.78 (.08)</u>	<u>4.31 (.12)</u>
Contour different	4.24 (.08)	4.27 (.06)	3.91 (.08)	3.80 (.07)

Table 5

Experiment 1 regression equation.

	Standardised Beta	*t*	*p*	Zero-order correlation	sr^2	Tolerance
Intercept		37.125	.000			
Rhythm	.570	12.934	.000	.613	.268	.823
Contour	.461	11.522	.000	.462	.212	1.000
Metricality	-.100	-2.266	.024	-.339	.008	.823

Total r^2 = .597

Table 6

Factor scores from principal components analysis of Experiment 1 ratings (plotted in Figures 5-6). Columns are sorted in order of variance accounted for; labels are post-hoc interpretations. Note rows are sorted first by time level.

Variant	Rhythm	Contour	Metre	Tonality
p1t1	-1.24	-0.65	-1.01	-0.30
p2t1	-1.06	-0.04	0.59	0.04
p3t1	-1.11	1.30	-0.44	0.07
p4t1	-0.82	0.94	0.16	0.58
p1t2	-1.27	-1.23	0.53	-0.50
p2t2	-0.73	-0.76	1.50	0.51
p3t2	-0.60	1.06	0.33	-0.83
p4t2	-0.40	0.87	0.21	0.83
p1t3	0.15	-1.17	-1.81	0.14
p2t3	0.76	-1.10	-0.91	2.25
p3t3	0.75	0.61	-1.46	-1.57
p4t3	0.98	0.86	-1.08	0.49
p1t4	0.66	-1.37	0.52	-1.13
p2t4	1.39	-0.89	1.25	0.03
p3t4	1.20	0.33	0.52	-1.61
p4t4	1.36	1.25	1.10	1.02

Table 7

Correlations between distances calculated using factor scores (see text) and regression predictors for Experiment 1. Columns are ordered by the percent variance accounted for by the assigned factor, as determined by which predictor had the highest correlation with each factor (bolded diagonal values).

	Rhythm	Contour	Metre	Tonality
Factor 1 (39% variance)	**-0.78**[a]	-0.04	0.32[a]	0.10
Factor 2 (28% variance)	-0.06	**-0.74**[a]	0.04	0.07
Factor 3 (13% variance)	-0.37[a]	-0.02	**0.41**[a]	0.19[b]
Factor 4 (7% variance)	-0.18[b]	-0.23[a]	0.05	**0.42**[a]

[a] denotes $p < .001$, [b] denotes $p < .01$

Table 8

Experiment 2 regression equation.

	Standardised Beta	*t*	*p*	Zero-order correlation	sr^2	Tolerance
Intercept		34.935	.000			
Contour	.413	7.791	.000	.450	0.161	.944
Rhythm	.320	6.214	.000	.329	0.102	.997
Tonality	-.155	-2.914	.004	-.268	0.023	.942

Total r^2 = .350

Table 9

Factor scores from principal components analysis of Experiment 2 ratings (plotted in Figures 11-12). Columns are sorted in order of variance accounted for; labels are post-hoc interpretations. Note this order diverges from Experiment 1, and that rows are accordingly sorted first by pitch levels.

Variant	Contour	Tonality	Rhythm	Factor 4	Factor 5
p1t1	-1.41	-0.09	0.38	-1.55	-0.35
p1t2	-1.83	-0.20	-0.38	-0.12	0.00
p1t3	-0.63	-0.33	0.60	0.01	-2.19
p1t4	-0.81	-0.21	0.82	-0.69	0.71
p2t1	-0.95	0.90	-0.90	1.07	0.54
p2t2	-1.01	0.96	-0.23	1.53	0.63
p2t3	-0.06	1.20	1.34	0.19	0.58
p2t4	1.14	0.41	0.79	1.96	0.10
p3t1	0.40	-2.09	-0.83	-0.28	1.20
p3t2	0.52	-1.96	0.23	0.69	0.85
p3t3	0.47	-0.51	0.92	-0.14	0.43
p3t4	0.65	-0.74	0.58	0.51	-2.30
p4t1	0.01	0.02	-2.42	-0.20	-0.55
p4t2	1.44	0.77	-1.16	-0.22	-0.36
p4t3	1.19	1.12	-0.65	-1.18	0.02
p4t4	0.91	0.76	0.93	-1.56	0.69

Table 10

Correlations between factor scores (listed with percent variance accounted for) and regression predictors for Experiment 2. Note ordering the columns by percent variance accounted for with the assigned factor (as in Table 7) yields a different order of predictors. Factors 4 and 5 have no obvious matching predictor.

	Contour	Tonality	Rhythm	Metre
Factor 1 (28% variance)	**-.438**[a]	.165[b]	-.074	.102
Factor 2 (19% variance)	-.247[a]	**.507**[a]	-.016	.048
Factor 3 (14% variance)	-.221[a]	.119	**-.309**[a]	.235[a]
Factor 4 (9% variance)	-.153[c]	.229[a]	-.094	.119
Factor 5 (7% variance)	-.029	.051	-.124[c]	.031

[a] denotes $p < .001$, [b] denotes $p < .01$, [c] denotes $p < .05$

Figure Captions

Figure 1. Example melody variants for Experiment 1. Sequences were always transposed within a given trial, such that they would start on different notes, but for ease of comparison are not transposed here.

Figure 2. Plot of Experiment 1 similarity ratings. The ascending diagonal represents match conditions (e.g., p3t2-p3t2), and accordingly has the highest similarity ratings.

Figure 3. Perceived similarity of all pitch manipulation levels in Experiment 1, averaged across time manipulation levels. Note the resulting change in colour scale from Figure 2.

Figure 4. Perceived similarity of all time manipulation levels in Experiment 1, averaged across pitch manipulation levels.

Figure 5. Factors 1 and 3 (interpreted as rhythmic pattern and metre) of the factor analysis solution of perceived similarity ratings of all compared variants, for Experiment 1. For clarity, data labels emphasise time levels, and internal axes crossing at the origin are added.

Figure 6. Factors 2 and 4 (interpreted as contour and tonality) of the factor analysis solution of perceived similarity ratings of all compared variants, for Experiment 1. For clarity, data labels emphasise pitch levels, and internal axes crossing at the origin are added.

Figure 7. Example variants used in Experiment 2. The stimuli were the same as Experiment 1 (see Figure 1) except for the added chord cadence prefix and variable tempo (across melody).

Figure 8. Plot of Experiment 2 similarity ratings. The ascending diagonal represents match conditions, which have the highest similarity ratings. The axes have been reordered from the Experiment 1 data (Figure 2) in accordance with the relative explanatory value of dimensions in similarity ratings.

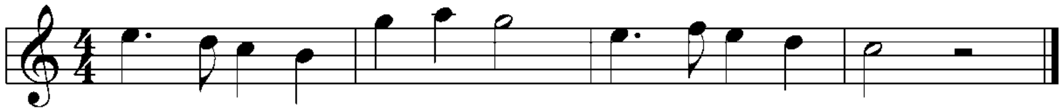
Figure 9. Perceived similarity of all pitch manipulation levels in Experiment 2, averaged across time manipulation levels.

Figure 10. Perceived similarity of all time manipulation levels in Experiment 2, averaged across pitch manipulation levels.

Figure 11. Factors 1 and 2 (interpreted as contour and tonality) of the factor analysis solution of perceived similarity ratings of all compared variants, for Experiment 2. For clarity, data labels emphasise pitch levels, and internal axes crossing at the origin are added.

Figure 12. Factors 3 (interpreted as metre) and 4 (uninterpreted) of the factor analysis solution of perceived similarity ratings of all compared variants, for Experiment 2. For clarity, data labels emphasise time levels, and internal axes crossing at the origin are added.

(a) p1t1: Pitch original, Time original



(b) p1t2: Pitch original, Ametric original rhythm

(c) p1t3: Pitch original, Rhythm-violated

(d) p2t1: Atonal original contour, Time original

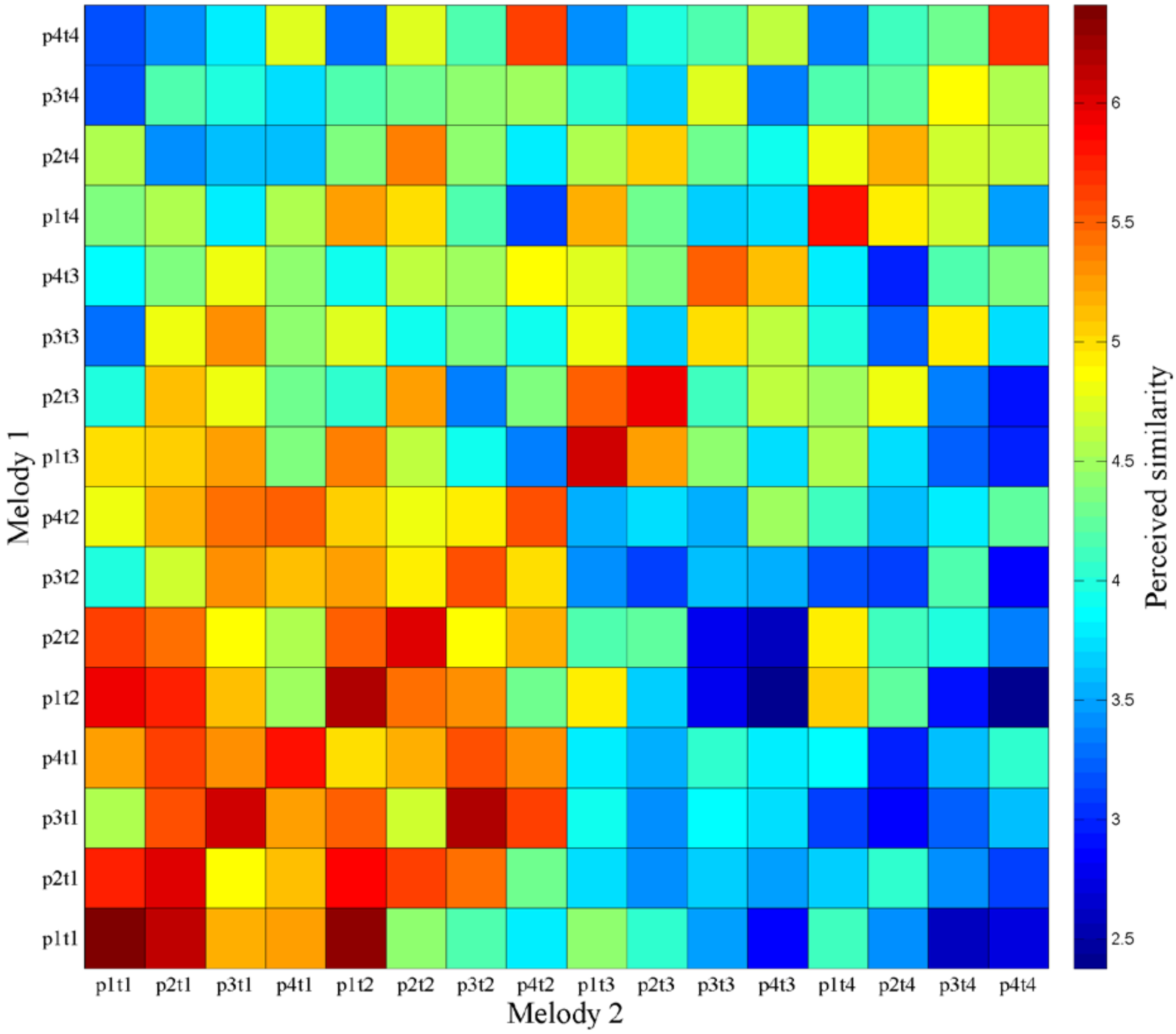
(e) p2t3: Atonal original contour, Rhythm-violated

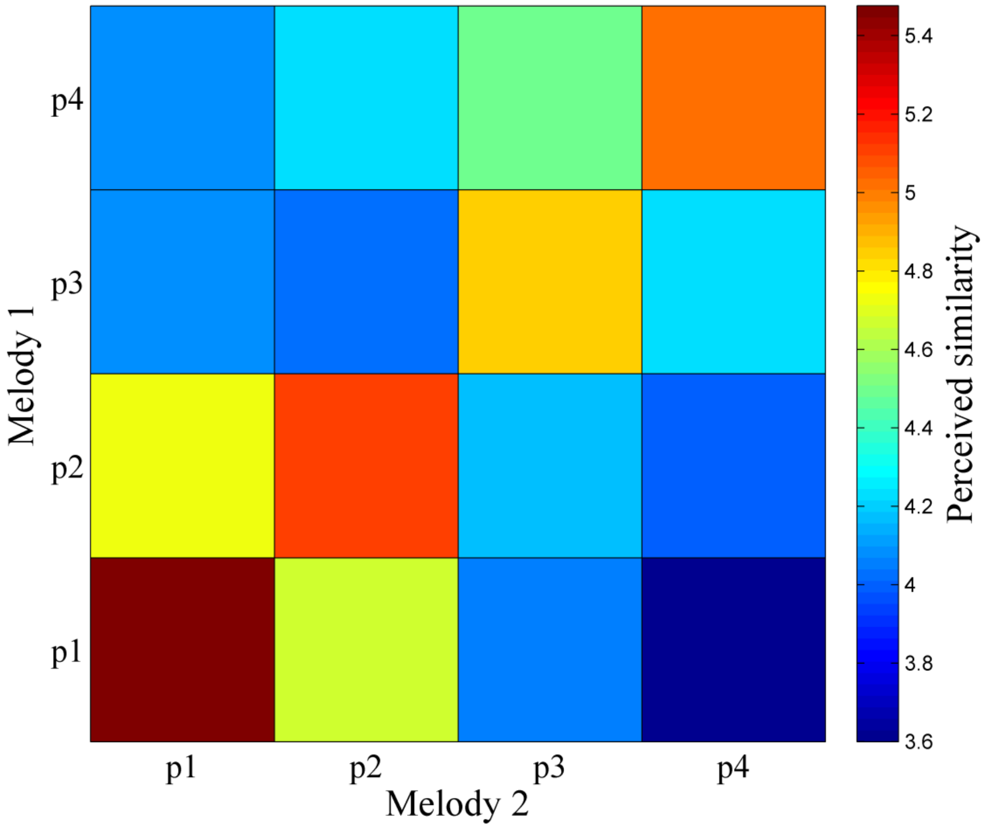
(f) p3t1: Contour-violated, Time original

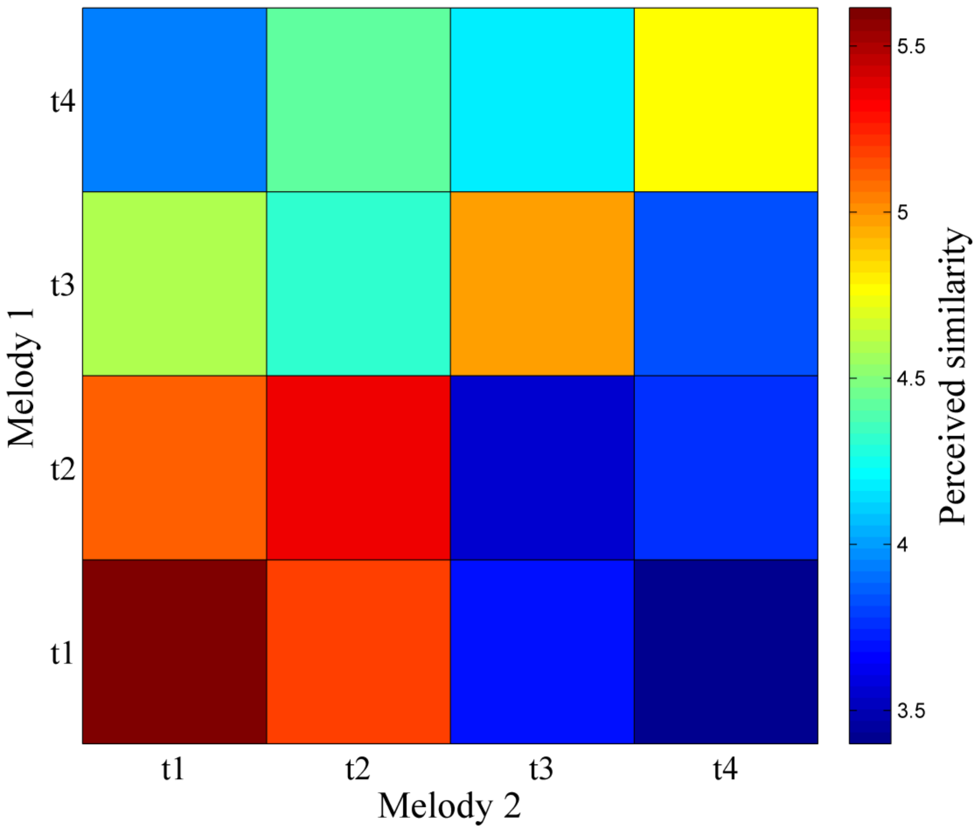
(g) p3t2: Contour-violated, Ametric original rhythm

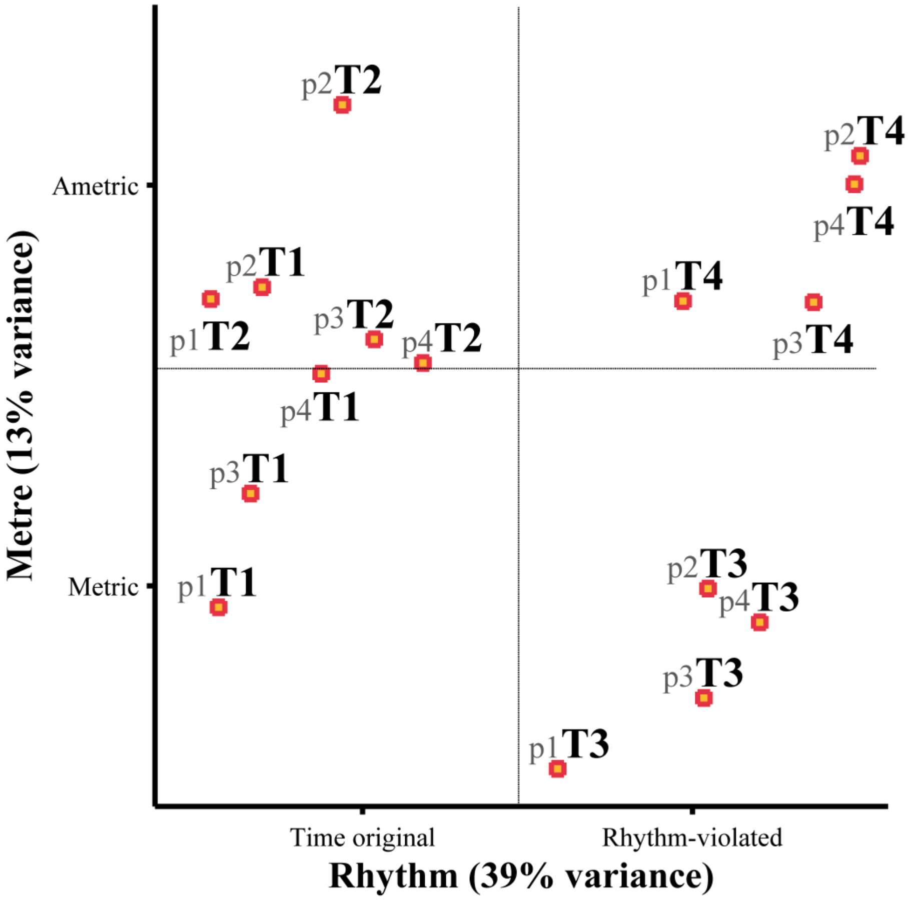
(h) p4t4: Contour-violated-atonal, Rhythm-violated-ametric

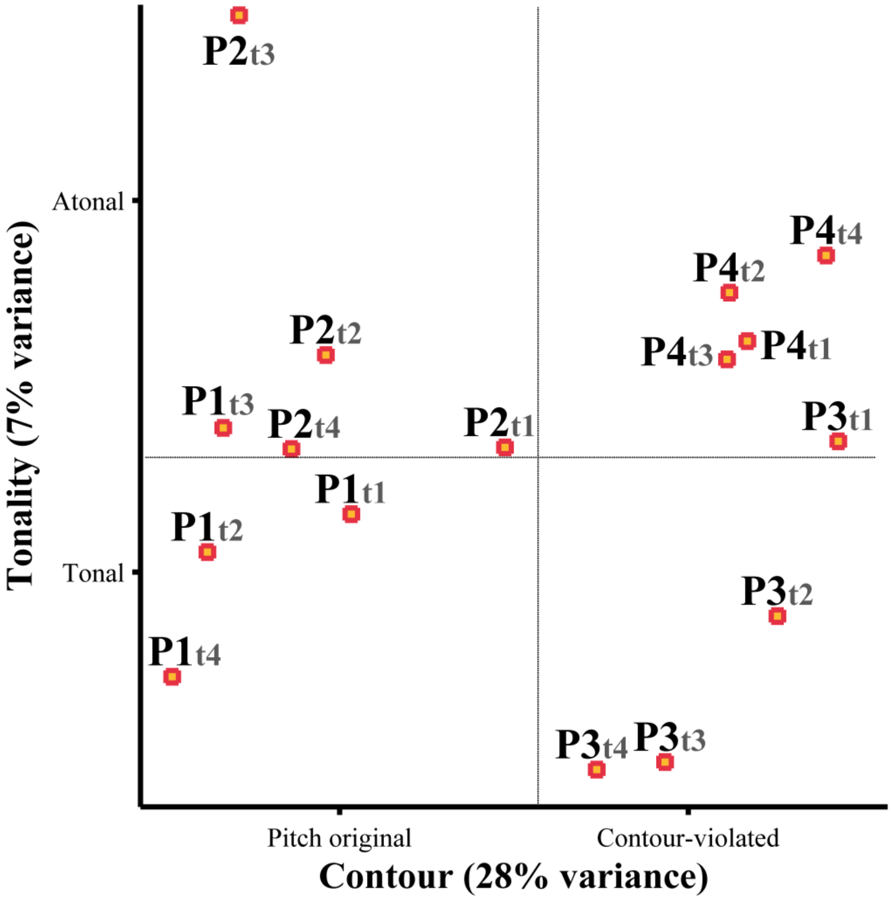

3
3


p4t4
p3t4
p2t4
p1t4
p4t3
p3t3
p2t3
p1t3
p4t2
p3t2
p2t2
p1t2
p4t1
p3t1
p2t1
p1t1
Melody 1
p1t1 p2t1 p3t1 p4t1 p1t2 p2t2 p3t2 p4t2 p1t3 p2t3 p3t3 p4t3 p1t4 p2t4 p3t4 p4t4
Melody 2
6
5.5
5
4.5
4
3.5
3
2.5
Perceived similarity


p4
p3
p2
p1
Melody 1
p1
p2
p3
p4
Melody 2
5.4
5.2
5
4.8
4.6
4.4
4.2
4
3.8
3.6
Perceived similarity

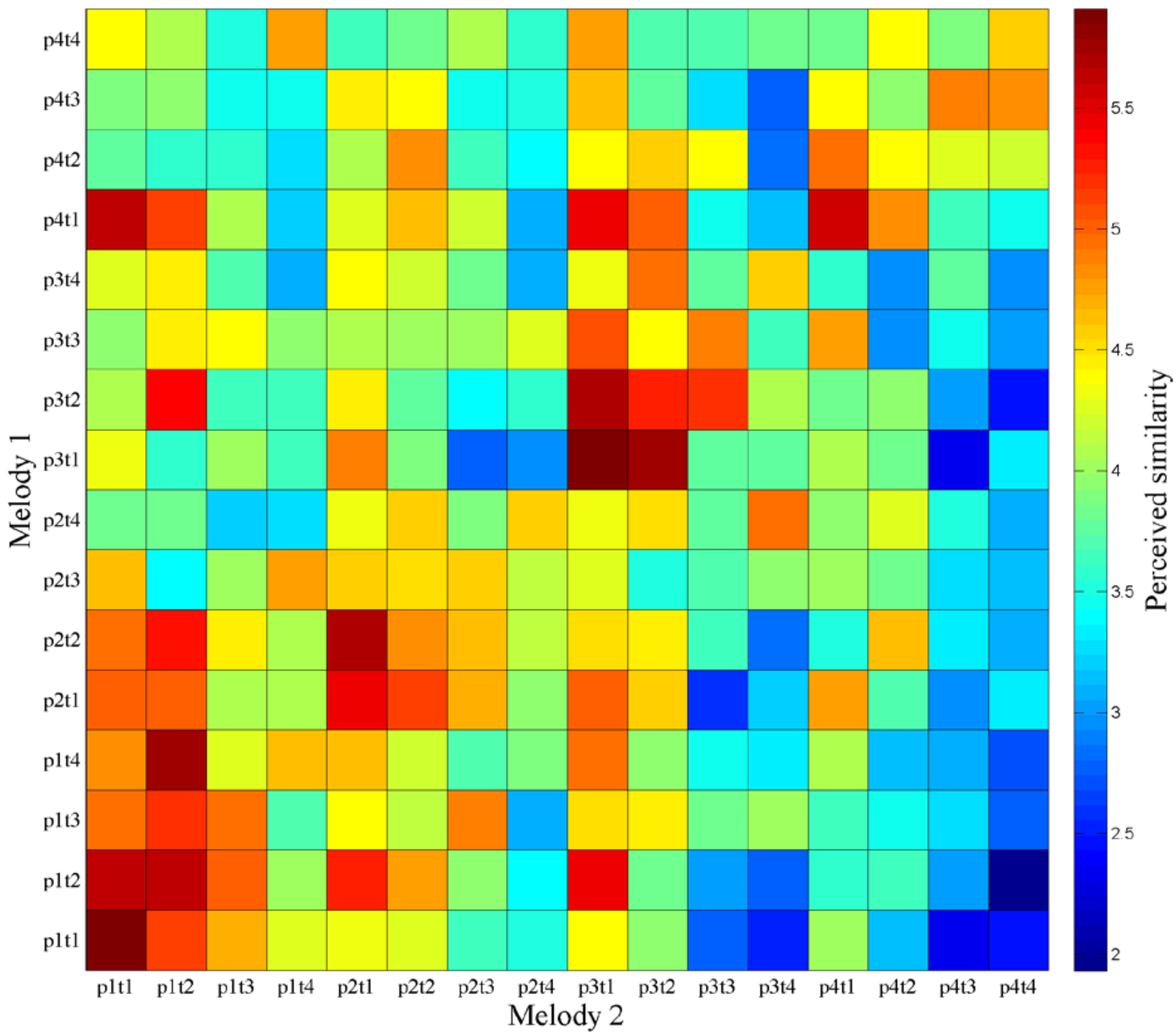

t4
t3
t2
t1
Melody 1
t1
t2
t3
t4
Melody 2
5.5
5
4.5
4
3.5
Perceived similarity

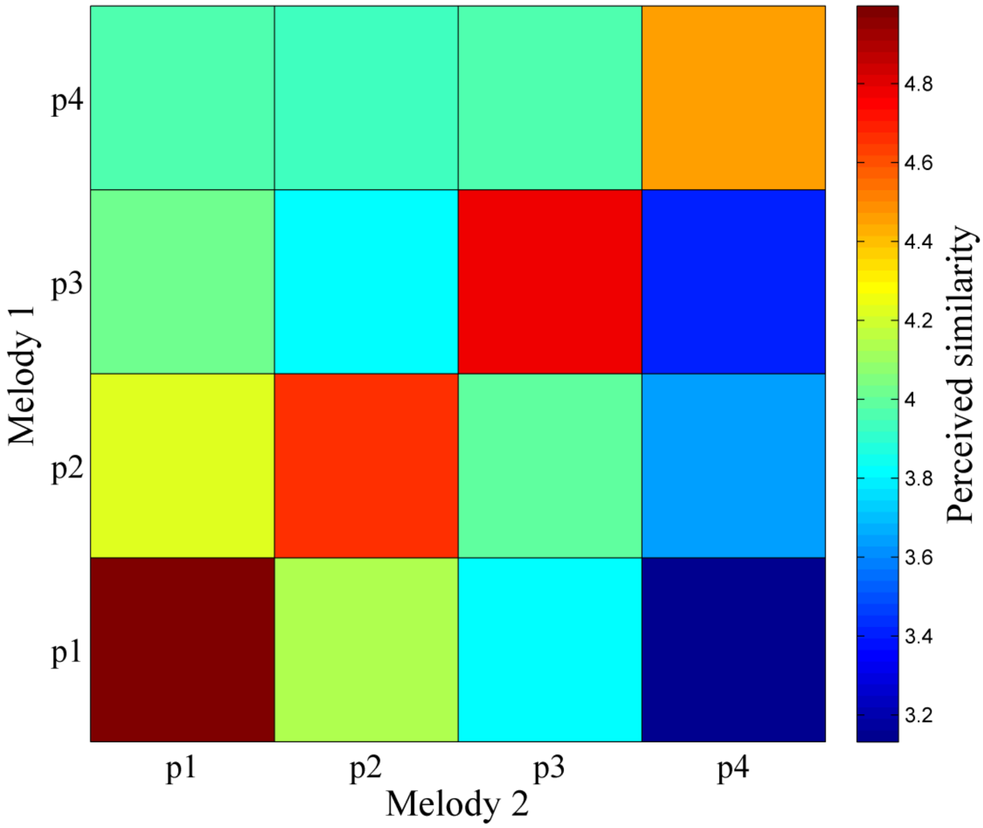

p2T2
p2T4
Ametric
p4T4
p2T1
p1T4
p3T2
p1T2
p4T2
p3T4
p4T1
p3T1
p2T3
Metric
p1T1
p4T3
p3T3
p1T3
Time original
Rhythm-violated
Rhythm (39% variance)
Metre (13% variance)

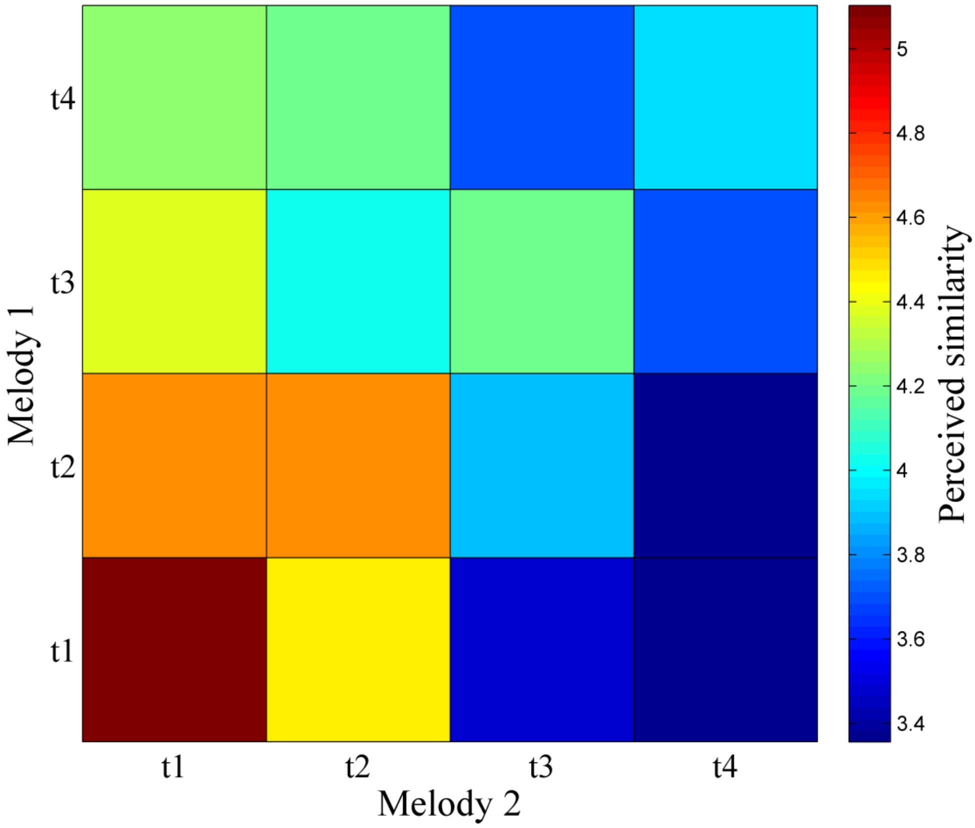

P2t3
Atonal
P4t4
P4t2
P2t2
P4t3
P4t1
P1t3
P2t4
P2t1
P3t1
P1t1
P1t2
Tonal
P3t2
P1t4
P3t4
P3t3
Tonality (7% variance)
Pitch original
Contour-violated
Contour (28% variance)

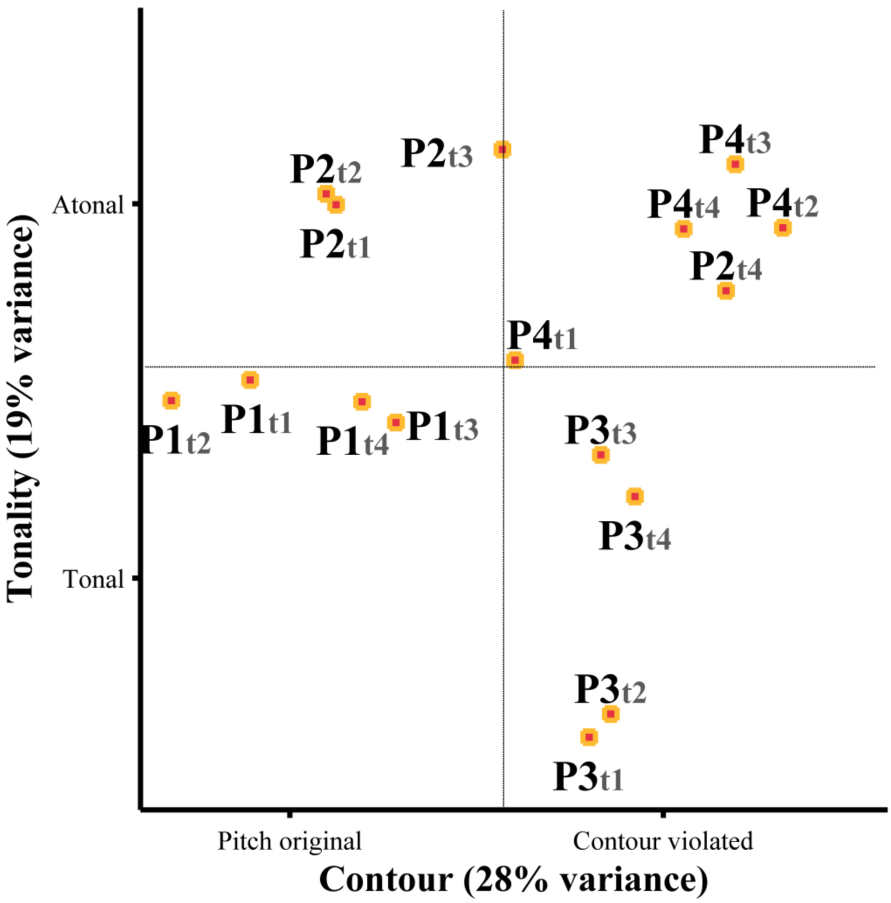
(a) p1t1: Pitch original, Time original

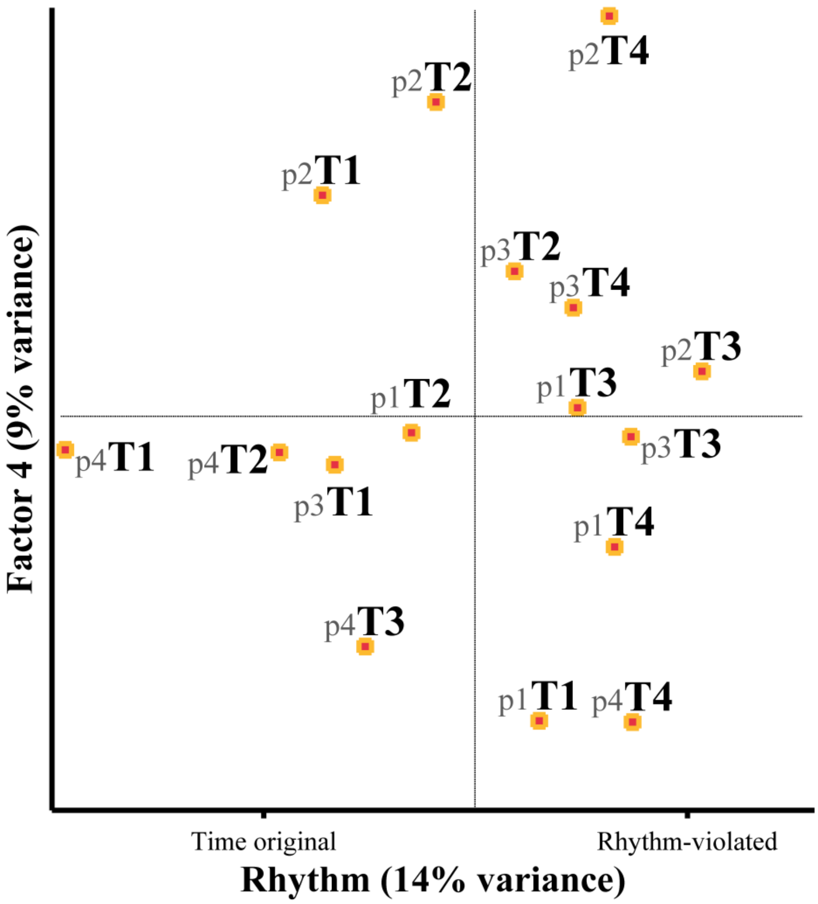
(b) p2t3: Atonal original contour, Rhythm-violated


p4t4
p4t3
p4t2
p4t1
p3t4
p3t3
p3t2
p3t1
p2t4
p2t3
p2t2
p2t1
p1t4
p1t3
p1t2
p1t1
Melody 1
p1t1 p1t2 p1t3 p1t4 p2t1 p2t2 p2t3 p2t4 p3t1 p3t2 p3t3 p3t4 p4t1 p4t2 p4t3 p4t4
Melody 2
5.5
5
4.5
4
3.5
3
2.5
2
Perceived similarity


p4
p3
p2
p1
Melody 1
p1
p2
p3
p4
Melody 2
4.8
4.6
4.4
4.2
4
3.8
3.6
3.4
3.2
Perceived similarity


t4
t3
t2
t1
Melody 1
t1
t2
t3
t4
Melody 2
5
4.8
4.6
4.4
4.2
4
3.8
3.6
3.4
Perceived similarity


P2t3
P4t3
P2t2
Atonal
P4t4
P4t2
P2t1
P2t4
P4t1
P1t2
P1t1
P1t4
P1t3
P3t3
P3t4
Tonal
P3t2
P3t1
Tonality (19% variance)
Pitch original
Contour violated
Contour (28% variance)


p2T4
p2T2
p2T1
p3T2
p3T4
p2T3
p1T3
p1T2
p3T3
p4T1
p4T2
p3T1
p1T4
p4T3
p1T1
p4T4
Factor 4 (9% variance)
Time original
Rhythm-violated
Rhythm (14% variance)