

THE DISTINCTIVENESS EFFECT IN THE RECOGNITION OF WHOLE MELODIES

MIRIAM RAINSFORD, MATTHEW A. PALMER, &
JAMES D. SAUER
University of Tasmania, Tasmania, Australia

DISTINCTIVE STIMULI ARE BETTER RECOGNIZED than typical stimuli in many domains (e.g., faces, words). Distinctiveness predicts the point of recognition of a melody (Bailes, 2010), and the recognition of unique tones within a melody (Vuvan, Podolak, & Schmuckler, 2014), yet no studies have examined the role of distinctiveness in recognizing whole melodies. We composed a set of novel melodies according to rules that should result in these being perceived as more or less distinctive. Using computational analysis and human ratings by a group of 36 pilot testers, we established a final stimulus set of 96 novel melodies (48 eight-note, 48 sixteen-note), half of which were high and half low in distinctiveness. A separate group of 26 participants completed a recognition test using this stimulus set. Using linear mixed-effects modeling, we found that greater pitch and interval range, wider intervals, varied contour, and ambiguous tonality within a Western diatonic framework predicted human perception of distinctiveness. However, only a wider modal (most frequent) interval predicted correct recognition. Distinctiveness improved recognition performance in both stimulus lengths; however, a significant advantage was only shown for sixteen-note melodies. Thus, the distinctiveness effect as observed across domains generalizes to the recognition of longer, whole melodies.

Received: January 31, 2017, accepted November 2, 2018.

Key words: distinctiveness effect, melody, recognition, mirror effect, computational modeling

THE DISTINCTIVENESS OF AN ITEM REFERS TO the degree to which the item possesses unusual or unique features (Schacter & Wiseman, 2006). Across domains, distinctive items are better recognized than those that are more prototypical (e.g., recognition of forenames, Brandt, Gardiner, & McCrae, 2006; word recognition, Israel & Schacter, 1997; Schacter & Wiseman, 2006; facial recognition, Valentine, 1991). The distinctiveness effect in memory, first proposed by von

Restorff (1933), has been replicated in visual (Bülthoff & Newell, 2015; Cohen & Carr, 1975) and verbal recognition (Dewhurst & Parry, 2010; Kausler & Pavur, 1974; Rajaram, 1998). Distinctiveness not only improves correct recognition, but reduces false alarms and false memory effects (Israel & Schacter, 1997; Schacter & Wiseman, 2006). However, few studies have investigated the role of distinctiveness in the recognition of musical material. Distinctiveness has been identified as a factor in specific aspects of music recognition, including improved encoding of melodic material in comparison to rhythmic patterns (Hébert & Peretz, 1997), facilitation of the point of recognition at which a listener can identify a melody (Bailes, 2010), and recognition of unique tones within a melody (Vuvan, Podolak, & Schmuckler, 2014). However, no studies have examined the distinctiveness effect when recognizing whole melodies. Therefore, we tested whether a set of melodies constructed to be high in distinctive features would be better recognized by participants in an old-new recognition test than a group of melodies constructed and measured to be of low distinctiveness.

Studies that have investigated distinctiveness in music have identified certain features of a melody that may cause it to be perceived as more or less distinctive. One such feature is the perceived tonal distance between musical keys. Western music theory defines a hierarchical series of tonal relationships around a central pitch or *tonic* note (Krumhansl, 1991). Krumhansl and Kessler (1982) demonstrated the psychological perception of this hierarchy by using a probe tone technique, where listeners judged how well a probe tone followed a musical event such as a scale or chord. These ratings were used to derive a set of key profiles of the relative distance between tones of the chromatic scale and the Western major and minor scales, showing that perceived tonal distance conformed to the structures used in music theory. Tones relating to the diatonic major and minor chords, such as the tonic, third, and fifth of the scale, were judged as better fitting in comparison to the minor second, augmented fourth, and seventh. These key profiles were further used to generate a map of perceived tonal distances between Western musical keys.

Tonal distance is related to distinctiveness; events close to the tonic are more predictable, whereas events that are peripheral are perceived as unexpected or

distinctive (Schmuckler, 1997). Vuvan and colleagues (2014) tested recognition of single target tones within a melody, finding an advantage in recognition for tones with greater distance from the tonic note of the scale according to Krumhansl and Kessler's (1982) profiles. These highly unexpected, schema-incongruent tones represented distinctive events within a tonal melody, and were thus better recognized. In the same study, however, highly expectable, tonally congruent probe tones were also better recognized, although this was explained as consistent with an availability heuristic, suggesting that these two processes operate simultaneously in the recognition of tonal information (Vuvan et al., 2014).

Further, certain scale degrees or intervals may also be perceived as distinctive, depending on their relationship to the tonic. Bailes (2010) examined the role of distinctiveness in the point of recognition (POR) at which a listener can name an earlier-heard melody. Using the Humdrum toolkit (Huron, 1993), Bailes (2010) calculated the probability of occurrence of intervallic and scale-degree information within a large corpus of German folk melodies. Consistent with Krumhansl and Kessler's (1982) key profiles, scale degrees peripheral to the scale, such as the tritone, augmented sixth (in major scales), and augmented seventh, were less probable within a melody than notes tonally close to the tonic such as the second, third, and fifth degrees of the scale. In addition, wide intervallic leaps such as a descending augmented fourth were found to be less probable, and thus more distinctive, than stepwise motion such as the descending minor second. In a gating paradigm experiment where melodies were presented note-by-note, melodies with a high content of low probability, and thus, distinctive scale-degree and intervallic events, were associated with an earlier POR at which the melody could be identified by the participant as previously presented, with 84.9% of the variance in POR explained by the level of distinctive information contained in its melodic, rhythmic, and scalar features (Bailes, 2010).

Müllensiefen and Halpern (2014) also found evidence that would support an advantage for distinctive over typical stimuli when recognizing whole melodies. Following an old-new recognition test of popular melodies, computer-based analysis using the software FANTASTIC (Müllensiefen, 2009a) was used to identify those features that elicited greater accuracy in performance. Correct recognition of old melodies was associated with infrequently used motifs in relation to the test set, and with a varied contour of wide intervallic leaps, features similar to those found by Bailes (2010) to be less probable, and thus distinctive. In contrast, melodies associated with

increased misses (failure to recognize an item as old) were found to have flat, stepwise contours, with motifs commonly used across the test set of stimuli.

However, in the same study, distinctive features did not predict correct rejection of lure melodies (Müllensiefen & Halpern, 2014). According to the mirror effect (Glanzer & Adams, 1985), the same features that predict improved recognition of old items should also be associated with correct rejection of lures. Thus, distinctive information should both improve hits (correct identification of targets) and reduce false alarms (incorrect identification of lures) in an old-new recognition test (Schacter & Wiseman, 2006). The mirror effect has been demonstrated in the recognition of distinctive faces (Cohen & Carr, 1975), low frequency words (Pazzaglia, Staub, & Rotello, 2014), and the recognition of words when accompanied by distinctive visual or auditory material (Dodson & Schacter, 2001; Israel & Schacter, 1997; Schacter, Cendan, Dodson, & Clifford, 2001). In contrast to these findings, Müllensiefen and Halpern (2014) identified that infrequently used, and thus distinctive melodic motifs, contributed both to correct identifications and false recognition of melodies. This was attributed to increased attention to distinctive features in both targets (old melodies) and lures (new melodies) as specially occurring events. If an interval or feature is registered as a special event during the test phase, it might trigger memory for similar motifs in old melodies, leading to mistaken recognition of the novel item as old (Müllensiefen & Halpern, 2014). However, Müllensiefen and Halpern (2014) did not manipulate distinctiveness directly, but instead used post hoc analysis of melodic features to identify a set of characteristics that predicted correct and false recognition.

The computational techniques used by Müllensiefen and Halpern (2014) to identify melodic features associated with improved accuracy in recognition might then be useful when creating a novel set of stimuli for the purposes of testing the distinctiveness effect in music. In the present experiment, we used FANTASTIC (Müllensiefen, 2009a) to obtain measurements of melodic features for a collection of novel eight- and sixteen-note melodic stimuli. This software combines approaches from descriptive statistics, music cognition, and computational linguistics to produce a series of measurements describing the features of a melody (Müllensiefen, 2009b). FANTASTIC computes *feature summary statistics* describing the content of the melody, including information about pitch and interval content, tonality, and contour, as well as calculations based on the frequency of occurrence of *m-types*, or short subsegments of the melody, similar to the concept of *n-grams* in

linguistics (Müllensiefen, 2009b). This software allowed us to identify specific melodic features associated with the perception of distinctiveness, as well as those features that are both perceived as distinctive and contribute to improved performance in a recognition test.

Rationale and Aim of the Present Study

Although distinctiveness has been identified as contributing to the recognition of individual tones (Vuvan et al., 2014) and the point of recognition of a melody (Bailes, 2010), it is surprising that no studies have investigated whether the distinctiveness effect, as observed across other domains (Schacter & Wiseman, 2006), generalizes to melodic recognition. In this study, we investigated whether the distinctiveness effect facilitates improved recognition of whole melodies.

To do so, we conducted two experiments. First, we developed a set of novel stimuli for testing the distinctiveness effect in melodic recognition. We used Bailes' (2010) measurements of scale degree and interval probability as rules for creating a set of melodies of high distinctiveness (featuring many low probability events) and melodies of low distinctiveness (featuring many high probability events).

We obtained subjective ratings of perceived distinctiveness of these melodies from a group of pilot testers. We also submitted the melodies to computational analysis using the software FANTASTIC (Müllensiefen, 2009a) to identify musical features of pitch, contour, interval, and tonality that were associated with human perception of distinctiveness in whole melodies. Following these analyses, we further identified two subgroups of melodies containing the most and least distinctive features for use in testing for the distinctiveness effect in melodic recognition. We further verified, using FANTASTIC, that the two subgroups differed substantially on those properties identified as associated with distinctiveness.

In a second experiment, we conducted an old-new recognition test using this final set of stimuli, to investigate whether distinctiveness improves recognition of whole melodies. Testing was conducted in two blocks, one for each stimulus length (eight- and sixteen-note). In each block of trials, participants first listened to a counterbalanced selection of 24 melodies, half of which were from the high distinctiveness subgroup, and half from the low distinctiveness subgroup. Participants were then tested for recognition of these melodies within the full corpus of 48 melodies. We predicted, following Bailes (2010) and Müllensiefen and Halpern (2014), that in both tasks performance would be greater

for melodies rated as being highly distinctive than those rated as highly typical. Using the melodic features measured in the previous experiment, we further used computer-based modeling to investigate those features which contributed to improved recognition.

Experiment 1

We composed a group of 156 stimuli (78 eight-note, 78 sixteen-note) according to rules that should allow them to be perceived as more or less distinctive. We then obtained participant ratings of distinctiveness for the full group of melodies. We submitted the final selection of melodies (96 melodies; 48 eight-note and 48 sixteen-note) for analysis using FANTASTIC, and compared these results with participant ratings to identify those musical features which contributed to the perception of distinctiveness.

METHOD

PARTICIPANTS

The sample consisted of 36 international, English-speaking participants who were recruited to take part in an online experiment. Demographic information was not collected due to experimenter error.

STIMULUS CREATION AND SELECTION

Stimuli consisted of 96 melodies, 48 of which were eight notes long, and 48 sixteen notes long. These were selected from a larger corpus of 156 melodies that were composed according to the rules specified below. For each stimulus length, 24 melodies of high and 24 of low distinctiveness were included. All melodies were composed to an isochronous rhythm of quarter notes to hold rhythmic factors constant (Hébert & Peretz, 1997). Stimuli were composed on a modal scale commonly used in world musics (*Maqam Kurd*, in Arabic music, also known as the *Phrygian mode* in medieval music), in order to reduce the likelihood that the stimuli might cue a similar, familiar melody in memory (Sloboda & Parker, 1985).

Compositional rules. Stimuli were composed according to two measures used by Bailes (2010) to determine the relative level of distinctiveness of a melody: intervallic probability and scale degree probabilities. Bailes (2010) used the Humdrum toolkit (Huron, 1993) to compute a series of bit rates indicating the relative probability of occurrence of each interval of the diatonic scale, and each scale degree within the Western major and minor scales. Wider intervallic material was found to have a lower frequency of occurrence (Bailes, 2010),

a finding supported by Müllensiefen and Halpern's (2014) analysis of the features of distinctive melodies.

Although the data for scale degrees presented by Bailes (2010) were computed for Western major and minor scales, and not the Phrygian mode used in our stimuli, participants with a Western listening background would be likely to perceive melodies in terms of Western constructs of consonance and dissonance acquired through passive listening experiences (Johnson-Laird, Kang, & Leong, 2012). Chords are perceived as consonant when they are consistent with schemata in long term memory, acquired via priming through music listening. Dissonance is perceived when a chord is incongruent with these schemata; dissonance ratings decrease as a listener becomes more familiar with pitch combinations. Thus, the perception of consonance is a cognitive process, learned through repeated exposure (McLachlan, Marco, Light, & Wilson, 2013). Tonal schemata are specific to one's cultural and music listening background, and are shown to be acquired implicitly in both trained and untrained musicians (Krumhansl, 1991; Krumhansl & Kessler, 1982; Stevens & Byron, 2009). Therefore, in composing our stimuli we used the bit rate data obtained by Bailes (2010) for major and minor scales with the expectation that a Western listener would perceive a novel melody within the context of already acquired schemata (Krumhansl, 1991; McLachlan et al., 2013; Vuvan et al., 2014).

The data obtained by Bailes (2010) were used as compositional rules to create the stimulus set; when composing melodies of high distinctiveness, wider intervallic leaps and less-frequently used notes of the scale were included, whereas melodies of low distinctiveness comprised commonly used notes of the scale, with flat, stepwise contours. Figure 1 shows an example of a high distinctiveness (panel A) and a low distinctiveness (panel B) sixteen-note melody composed according to these rules.

COMPUTATIONAL ANALYSIS OF STIMULUS FEATURES

Summary of features analyzed. FANTASTIC is capable of computing both first- and second-order features of a melody. First-order features are calculated based on the content of the melody itself (Jakubowski, Finkel, Stewart, & Müllensiefen, 2016). These include descriptive statistics, referred to in FANTASTIC as *feature value summary statistics*, as well as *m-type summary statistics*, which are calculated using m-types, or brief subsegments of the melody, similar to the concept of *n-grams* in linguistics (Müllensiefen, 2009b). Second-order features describe the frequency of occurrence of features relative to a corpus or collection of melodies



FIGURE 1. Samples of sixteen-note melodies from the high distinctiveness (panel A) and low distinctiveness (panel B) melody sets. The high distinctiveness melody features wide intervallic leaps, whereas the low distinctiveness melody features stepwise motion.

TABLE 1. *Glossary of FANTASTIC Variable Names*

Variable name	Meaning
p.range	Pitch range
p.entropy	Pitch entropy
p.std	Pitch standard deviation
i.abs.range	Interval absolute range
i.abs.mean	Interval absolute mean
i.abs.std	Interval absolute standard deviation
i.mode	Modal interval
i.entropy	Interval entropy
tonalness	Tonalness
tonal.clarity	Tonal clarity
tonal.spike	Tonal spike
int.cont.glob.dir	Interpolation contour global direction
int.cont.grad.mean	Interpolation contour mean gradient
int.cont.grad.std	Interpolation contour gradients standard deviation
int.cont.dir.changes	Interpolation contour direction changes
step.cont.glob.var	Step contour global variation
step.cont.glob.dir	Step contour global direction
step.cont.loc.var	Step contour local variation
poly.coeff1	Polynomial contour coefficient 1
poly.coeff2	Polynomial contour coefficient 2
poly.coeff3	Polynomial contour coefficient 3

(Jakubowski et al., 2016). Table 1 presents a glossary of the FANTASTIC variable names and their meaning.

In this study, we computed first-order feature value summary statistics and m-type summary statistics of the melodies, analyzing pitch, interval, contour, and tonality. We excluded all statistics describing rhythm as these were uniform across all melodies due to the isochronous rhythm of our stimuli. We were unable to include the second-order features analyzed by Müllensiefen and Halpern (2014), as these cannot presently be calculated using FANTASTIC in a corpus of melodies with isochronous rhythm because this causes the algorithms to divide by zero.

Pitch features analyzed included pitch range (from the lowest to highest note, in semitones), variance (standard deviation), and entropy. Entropy calculations in FANTASTIC are based on Shannon entropy in information theory (Shannon, 1948), and describe the relative frequency of events; thus, pitch entropy describes the frequency of occurrence of the pitch classes of a melody (Müllensiefen, 2009b). For intervallic features, we computed intervallic range and mean interval (in semitones), intervallic variance (standard deviation), and interval entropy (Müllensiefen, 2009b).

Three different methods were used to calculate the contour of melodies. *Interpolation contour* represents a melody as a series of straight lines interpolating between its extreme high and low points (Müllensiefen, 2009b). Using FANTASTIC, we calculated the global direction, mean, standard deviation, and number of changes in interpolation contour. *Step contour* represents a melody as a step curve, by plotting duration on the x-axis against pitch values on the y-axis (Müllensiefen, 2009b). We calculated the global direction and global and local variation in step contour. *Polynomial contour* represents the contour of a melody as a polynomial curve. Three coefficient statistics (*poly.coeff1*, *poly.coeff2*, and *poly.coeff3*) describe the three-dimensional variation in shape of the contour, thus capturing its major variations in direction (Müllensiefen, 2009b).

Analyses of the implicit tonality of the melodies in FANTASTIC is computed by using the Krumhansl-Schmuckler algorithm (Krumhansl, 1990) to calculate a *tonality vector* of length 24, consisting of the Pearson-Bravais correlation of the notes of the melody with all 24 Krumhansl-Kessler key profiles for the Western major and minor scales (Krumhansl & Kessler, 1982). *Tonalness* takes the highest value in the vector, thus, higher values indicate a stronger correlation with one of the major or minor scales. *Tonal clarity* is based on Temperley's (2007) statistic describing the degree of ambiguity in tonality. This is calculated from the ratio of the highest and second-highest correlation in the *tonality vector*. Higher values indicate closer correlations with a single key, rather than multiple keys, and are thus less ambiguous in tonality. *Tonal spike* further describes tonal ambiguity by dividing the highest correlation in the *tonality vector* by the sum of all correlation values which are greater than zero. Like *tonal clarity*, higher values indicate less ambiguity in tonality (Müllensiefen, 2009b).

PROCEDURE

Ratings. We obtained participant ratings of distinctiveness for the complete set of 156 melodies.

Participants were randomly assigned to complete a survey containing one of four randomly ordered presentations of the melodies. In each, participants first rated the eight-note stimuli, and then the sixteen-note stimuli, with twenty melodies per page. Melodies were presented as audio files in .wav format, using an HTML5 audio player. Participants were instructed to take a one minute break at the end of each page, and a five-minute break between the eight- and sixteen-note melodies.

For each melody, participants were asked to rate the statement "This melody has distinctive features" on a 7-point Likert-type scale ranging from -3 (*strongly disagree*) to +3 (*strongly agree*); 0 indicated *neither agree or disagree*. These ratings were then compared with the values obtained using computational analysis, to determine which melodic features were associated with human perception of distinctiveness.

Final stimulus set. To establish the final stimulus set for testing the effect of distinctiveness on melodic recognition, for each stimulus length we took the 24 melodies that had received the highest and lowest ratings of distinctiveness, for a total of 48 melodies in each stimulus length. The final selection of 96 melodies were submitted to computational analysis using the software FANTASTIC (Müllensiefen, 2009a), to verify that the high and low distinctiveness stimulus sets differed in musical properties relevant to the perception of distinctiveness.

Results and Discussion

We conducted two analyses. In the first, we examined which of the melodic features measured using FANTASTIC corresponded with participant ratings of distinctiveness. The second analysis established that the high and low distinctiveness melody collections differed substantially in the musical features identified in the first analysis, which would thus cause these melodies to be perceived as high or low in distinctiveness.

COMPARISON OF PARTICIPANT RATINGS WITH FEATURE ANALYSIS

To determine whether participant ratings of distinctiveness corresponded with an increase in melodic features that might be described as more or less distinctive, we conducted Bayesian correlations between the computed features of melodies and subjective ratings of distinctiveness. Orthodox Neyman Pearson statistical methods require an alpha correction for each comparison in a family of analyses to avoid false discovery, or Type I error. One problem with making such adjustments in a very large family of analyses is that adjustments may

TABLE 2. Bayesian Correlations Between Features of Melodies and Participant Ratings of Distinctiveness

Variable name	Distinctiveness (mean rating)	
	<i>r</i>	BF ₁₀
p.range	.49	51,289.56**
p.entropy	.32	19.46**
p.std	.51	133,359.80**
i.abs.range	.32	17.02**
i.abs.mean	.53	383,102.15**
i.abs.std	.31	12.02**
i.mode	.50	60,233.81**
i.entropy	.45	4,071.45**
tonalness	.36	80.38**
tonal.clarity	-.28	4.76*
tonal.spike	.01	0.13
int.cont.glob.dir	.03	0.13
int.cont.grad.mean	.10	0.21
int.cont.grad.std	.13	0.29
int.cont.dir.changes	-.04	0.14
step.cont.glob.var	.51	136,711.05**
step.cont.glob.dir	.01	0.13
step.cont.loc.var	.53	353,835.32**
poly.coeff1	-.16	0.42
poly.coeff2	-.08	0.17
poly.coeff3	.20	0.80

Note: * indicates substantial support for the hypothesis, ** indicates strong support for the hypothesis.

either be too liberal, and inflate Type I error, or too strict, and inflate the likelihood of missing a true effect, or Type II error (Curran-Everett, 2000). Bayesian procedures have recently been adopted in such situations as they are not affected by multiple comparisons (Dienes, 2011). Instead of setting an alpha level at which a result is considered significant, the Bayes factor BF_{10} calculates the likelihood that the observed data occurred under the alternative hypothesis H_1 rather than the null hypothesis H_0 . This is obtained from calculating the ratio of the posterior and prior odds of the alternate hypothesis being correct (Dienes, 2011). If $BF_{10} = 10$, the data are ten times more likely to have occurred under the alternate hypothesis than the null (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). Such procedures are therefore more appropriate for exploratory investigations, as they are not affected by whether the hypothesis was predicted before or after data collection (Dienes, 2011).

Table 2 presents Bayes factors and Pearson correlations between participant ratings of distinctiveness and the set of features calculated (see Table 1 for the FANTASTIC variable abbreviations and their meaning). According to Jeffreys' (1961) criteria, Bayes factors of 3 or above represent substantial evidence, and Bayes

factors of 10 or above represent strong evidence for the hypothesis that the variables were correlated.

Pitch and interval features. We found small-to-moderate positive correlations between distinctiveness and all pitch and intervallic features, representing strong support for the hypothesis. Thus, melodies that contained greater range and variability in pitch (p.entropy, p.std), wider intervals (int.abs.range, int.abs.mean, i.mode), and greater variability in the size of intervals used (i.entropy) were perceived as more distinctive by participants. Composition of the distinctive melodies by using more unusual notes of the scale and wider intervallic leaps—according to Bailes' (2010) computations of bit rate for Western major and minor scales—therefore resulted in an increase in pitch and intervallic variance, which corresponded with increased perception of melodies as distinctive by listeners.

Implicit tonality. A small-to-moderate positive correlation was also found between the tonalness of melodies and distinctiveness, representing strong support for the hypothesis. This was surprising, as it indicated that greater correspondence to a Western major or minor scale was perceived as distinctive, rather than typical. In the context of a corpus composed on a modal scale, it is possible that melodies corresponding more closely to Western scales may stand out, and thus appear more distinctive due to sensory priming of the modal scale on which our melodies were based (Bigand, Poulin, Tillmann, Madurell, & D'Adamo, 2003). In addition, for a sequence of wider pitches, intervals, or varied contour to be perceived as distinctive, this requires that these features are placed within a tonal context. Vuvan and colleagues (2014) found that highly unexpected, harmonically distant tones in relation to the Krumhansl and Kessler (1982) profiles were better recognized within a diatonic context, but within an atonal context, this advantage disappeared. We composed melodies intended to be distinctive using wider intervals as well as tonally distant notes of the scale. Where melodies were composed predominantly from notes peripheral to the tonic, it is possible that this caused the melodies to lose their distinctive nature.

However, a weak-to-moderate negative correlation between tonal.clarity and distinctiveness, representing substantial support for the hypothesis, indicated that as melodies became more ambiguous in key, they were perceived as more distinctive. A decrease in tonal clarity would have occurred via the use of less predictable notes of the musical scale when composing distinctive melodies. Although tonal.spike might also be expected to correlate with distinctiveness, examination

of the values obtained for our melodic corpus revealed a range of only 0.12 between the minimum value of 0.14 and the maximum of 0.26. This lack of variability within our stimuli may therefore have constrained our ability to obtain a meaningful correlation.

Melodic contour. The step contour of a melody is calculated from a vector drawn by plotting normalized duration values on the x-axis and pitch values on the y-axis. Global variation (`step.cont.glob.var`) refers to the standard deviation of the step contour vector, and describes the degree of variability in a melody's contour overall, whereas local variation (`step.cont.loc.var`) is calculated from the mean absolute difference between adjacent values of the vector, and thus reflects smaller scale changes in contour from note to note or within short motifs. Step contour global direction describes whether a melody descends or rises overall (Müllensiefen, 2009b).

We found strong, positive correlations between both `step.cont.glob.var` and `step.cont.loc.var` and perceived distinctiveness, representing strong support for the hypothesis. The wider intervallic leaps used in composition of the distinctive melodies would have resulted in greater variation in contour both at an overall (i.e., global) and local level; such variation was perceived as more distinctive. Step contour global direction (`step.cont.glob.dir`) did not correlate with distinctiveness; however, as this variable describes the overall direction of the melody, this feature would not be expected to be perceived as distinctive.

However, we obtained no support for correlations between distinctiveness and other measures describing contour. While it was surprising that global and local step contour were related to distinctiveness, but interpolation and polynomial contour were not, this may have been due to the brevity of the melodies. The four interpolation contour statistics describe the global variation (`int.cont.glob.dir`), mean gradient (`int.cont.grad.mean`), standard deviation of the gradient (`int.cont.grad.std`), and number of changes (`int.cont.dir.changes`) in interpolation contour. These statistics are derived from a series of gradients interpolating between the high and low points of a melody over set points in time (Müllensiefen, 2009b; Müllensiefen & Halpern, 2014). This method may not sufficiently capture variation in contour in our very brief melodies, thus, interpolation contour was not related to human perceptions of distinctiveness in our study, in contrast to Müllensiefen and Halpern's (2014) longer pop melodies. The three polynomial coefficients (`poly.coeff1`, `poly.coeff2`, `poly.coeff3`) represent the contour of a melody as the coefficients

of a polynomial curve. This statistic may again be better suited to longer and more complex melodies.

Summary of melodic features associated with the perception of distinctiveness. In summary, computational analysis of the stimulus set verified that the methods used to compose melodies high in distinctiveness resulted in an increase in specific features of pitch, interval, contour, and tonal complexity. Melodies perceived by participants as distinctive were associated with greater range and variability in pitch (`p.range`, `p.std`, `p.entropy`), wider intervals (`i.abs.range`, `i.abs.mean`, `i.mode`), and greater variability in the size of intervals used (`i.entropy`). Distinctive melodies corresponded more closely with Western diatonic scales (tonalness), but were also more ambiguous in key (`tonal.clarity`). Distinctive melodies contained a more varied contour overall (`step.cont.glob.var`), as well as increased changes in contour at a local level (`step.cont.loc.var`).

These features identified in our study as associated with the human perception of distinctiveness have some similarity to those that Müllensiefen and Halpern (2014) identified as contributing to improved recognition of old items. While their study did not directly manipulate distinctiveness, melodies that contained a highly varied contour (`int.cont.grad.std`) and unusual motifs in relation to the corpus of melodies tested (`mtcf.norm.log.dist.testset`, `mtcf.std.g.weight.testset`) were associated with improved recognition. Although their model identified a different measure of contour as predicting recognition (interpolation contour, cf. step contour in our study) this may be due to differences in stimulus type, as their study used longer melodic phrases taken from pop melodies. We were also unable to measure the m-type corpus features (`mtcf`) that they used, because these measures cannot currently be used with isochronous melodies, and so further comparison with their results is limited. However, composition of a melody with increased variety in pitch and intervallic content, as associated with distinctiveness in our study, would arguably result in a melody containing more unusual motifs. Further research is therefore needed to build a complete model of melodic and rhythmic features which predict distinctiveness.

Verification of high and low distinctive test sets. As a final measure, to ensure that the two groups of melodies differed sufficiently in musical features that would be perceived as distinctive, we conducted Bayesian independent samples *t*-tests to compare the two groups of melodies on all the computed features. We used the Cauchy prior of .707, which represents prior odds weighted slightly towards the null hypothesis (Rouder,

TABLE 3 Bayes Factor *t*-tests and Descriptive Statistics for High and Low Distinctiveness Melodies

Variable	BF ₁₀	Error %	Mean (SD)	
			High distinctiveness	Low distinctiveness
Distinctiveness	4.727e +34**	1.12e -41	0.81 (0.17)	0.14 (0.13)
p.range	3,157.10**	2.29e -9	9.57 (2.58)	6.85 (2.92)
p.entropy	12.89**	3.31e -6	0.47 (0.07)	0.42 (0.09)
p.std	3,035.37**	2.36e -9	3.52 (1.01)	2.50 (1.08)
i.abs.range	10.46**	4.93e -6	6.69 (3.08)	4.82 (3.00)
i.abs.mean	6,092.45**	1.43e -9	3.93 (1.66)	2.47 (1.14)
i.abs.std	6.87*	1.08e -5	2.53 (1.22)	1.83 (1.20)
i.mode	2,869.41**	2.46e -9	4.57 (1.72)	3.06 (1.34)
i.entropy	496.95**	7.47e -9	0.53 (0.07)	0.47 (0.08)
tonalness	11.78**	3.93e -6	0.69 (0.10)	0.63 (0.10)
tonal.clarity	12.77**	3.36e -6	1.14 (0.11)	1.22 (0.13)
tonal.spike	0.22	3.27e -4	0.19 (0.02)	0.19 (0.03)
int.cont.glob.dir	0.23	3.38e -4	-0.33 (0.86)	-0.40 (0.77)
int.cont.grad.mean	0.43	8.86e -7	3.23 (2.38)	2.63 (2.29)
int.cont.grad.std	0.60	7.23e -7	3.78 (3.33)	2.80 (2.94)
int.cont.dir.changes	0.22	3.24e -4	0.42 (0.37)	0.42 (0.39)
step.cont.glob.var	3,067.80**	2.35e -9	3.32 (0.95)	2.35 (1.02)
step.cont.glob.dir	0.23	3.36e -4	-0.05 (0.39)	-0.08 (0.39)
step.cont.loc.var	4,537.42**	1.77e -9	0.44 (0.19)	0.28 (0.13)
poly.coeff1	0.74	6.36e -7	-0.52 (2.94)	0.32 (1.86)
poly.coeff2	0.37	9.69e -7	-0.25 (3.43)	0.39 (2.16)
poly.coeff3	2.29	3.16e -7	0.18 (1.07)	-0.25 (0.70)

Note: * indicates substantial support for the hypothesis, ** indicates strong support for the hypothesis.

Speckman, Sun, Morey, & Iverson, 2009). The Cauchy prior is recommended when no previous research exists from which prior odds can be calculated (Dienes, 2011). As above, we chose to use Bayesian *t*-tests due to the risk of inflating Type II error when applying an alpha correction after such a large number of comparisons. Table 3 presents descriptive statistics and Bayes factors for the comparisons between high and low distinctiveness melody groups.

These analyses revealed that the high and low distinctiveness melody sets differed in human ratings of distinctiveness, as well as differing on the same set of features identified above as correlating with perceptions of distinctiveness. Strong support was obtained for greater variation in all measures of pitch and intervallic content in the high distinctiveness melodies than the low distinctiveness set, although the difference between groups was lesser for interval absolute standard deviation (*i.abs.std*), but still substantial according to Jeffrey's (1961) criteria. Strong support was also obtained that high and low distinctiveness melodies differed in measures of implicit tonality; as for the correlational analyses, the high distinctiveness melodies were more closely related to a Western diatonic scale (*tonalness*), but were more ambiguous in tonality (*tonal.clarity*). As for the correlational analyses, the high

distinctiveness melody collection contained greater variation in step contour at a global (*step.cont.glob.var*) as well as at a local level (*step.cont.loc.var*), in comparison to the low distinctiveness melodies. The two groups of melodies did not differ in tonal spike, interpolation contour measures, step contour global direction, and polynomial contour measures. However, as noted above, these properties were not observed to be related to perceived distinctiveness in this melody set. Thus, the analysis confirmed that the two melody sets differed in musical features that were also associated with perception of a melody as distinctive.

Experiment 2

In Experiment 1, we identified a set of musical features associated with perceived distinctiveness in brief melodies. Distinctive melodies contained greater variation in pitch and intervallic content, wider intervals, a more varied contour, and greater ambiguity in tonality within a Western diatonic framework. Thus, the use of compositional rules based on Bailes (2010) modeling of distinctive intervallic and scale-degree content resulted in melodies that were perceived as distinctive. We further established a subset of 96 melodies, half of which were high and half of which were low in distinctive features.

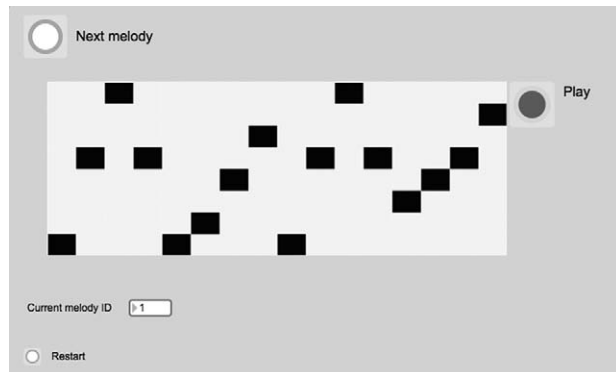


FIGURE 2. Participant view of stimulus presentation in the Exposure phase.

In Experiment 2, we used this subset of melodies in an old-new recognition test to examine whether melodies high in distinctive features would be better recognized than those low in distinctive features. Using mixed-effects models, we further investigated which of those melodic features identified in Experiment 1 as associated with perceived distinctiveness contributed to improved recognition performance.

Method

PARTICIPANTS

Participants were 29 first-year psychology students (4 male, 25 female) attending the University of Tasmania, who participated as part of a coursework requirement. Demographic information was not collected due to experimenter error. Data for three participants was lost due to a computer system failure. The final sample consisted of 26 participants (3 male, 23 female).

MATERIALS

We used the MUSOS Toolkit (Rainsford, Palmer, & Paine, 2018), a suite of programs developed in Max/MSP (Cycling '74, 2014) for the purpose of testing melodic recognition. During the exposure phase, stimuli were presented to participants using a *live.step* sequencer object, with the keyboard interface and beat grid removed, so that notes were displayed as black square blocks on an untextured, light grey background (see Figure 2). Participants used the Play button to listen to each melody once, and then used the Next Melody button to load the next melody. A piano sound (MIDI channel 1) was used for output so as to provide a pleasant but neutral timbre.

For the recognition test, the sequencer was removed and replaced with a progress bar to ensure that

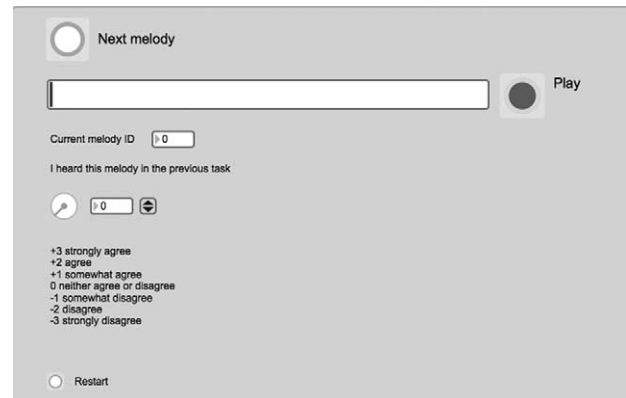


FIGURE 3. Participant view of interface for rating of melodies during Recognition testing. The sequencer object was removed and replaced with a progress bar.

participants did not use the visual display of melodies to cue recognition. Participants used the Play button to listen once to each melody, before rating whether they had heard the melody in the previous exposure phase, using either a dial or the up and down arrows to select the desired value (see Figure 3). After providing their rating, participants used the Next Melody button to load the next melody.

Participants completed two block of trials, one for the eight-note melodies, and one for the sixteen-note melodies. Participants were randomly assigned to complete either the eight-note or sixteen-note trials first. Each block of trials followed the same procedure, as follows.

PROCEDURE

Participants were given instructions by the experimenter on how to use the software to listen to melodies in the exposure phase, and how to listen to and respond to melodies in the recognition test. During these instructions no melodies were played. Participants were therefore aware that they were first required to study a series of melodies, and then complete a recognition test.

During the exposure phase, participants were first presented with 24 of the 48 melodies. Of these, 12 of were randomly selected from the database of 24 high distinctiveness melodies, and 12 randomly selected from the database of 24 low distinctiveness melodies. Participants were asked to listen carefully to each of the melodies. Following exposure, participants then completed a recognition test comprising all 48 melodies, including the 24 previously heard and 24 unheard melodies (12 high distinctiveness, 12 low distinctiveness) in random order. For each melody, participants rated their level of agreement with the statement “I heard this

melody in the previous task” on a 7-point scale ranging from -3 (*strongly disagree*) to +3 (*strongly agree*) with a midpoint of zero (*neither agree nor disagree*).

Results and Discussion

ROC CURVE ANALYSIS

Receiver operating characteristic (ROC) curve analysis was conducted on the participant ratings using pROC (Robin et al., 2011a). This method of analysis involves plotting the cumulative percentage of hits (HR) against false alarms (FAR) for each level of confidence in a decision, in order to obtain a measure of diagnostic accuracy in recognition as a continuous curve (Mickes, Flowe, & Wixted, 2012). The leftmost part of a ROC curve corresponds to decisions made with maximum confidence. These tend to be associated with very low false alarm rates and modest hit rates. Moving from left to right along the curve, the data include response made with cumulatively lower confidence ratings. Thus, moving from left to right, the hit rate and false alarm rate both increase. The right-hand extreme of the ROC curve corresponds to decisions made with all levels of confidence (from minimum to maximum); these decisions are characterized by very high hit rates (often close to 100%) and false alarm rates. From this plot, the area under the curve (AUC) is calculated. The diagonal line on the ROC curve plot spanning 0% to 100% has an AUC of 50%, indicating performance at chance level, and an AUC of 100% indicates perfect memory performance (Swets, 1973). When comparing two ROC curves, the curve

with the greatest AUC is therefore the most accurate (Mickes et al., 2012). Curves whose 95% confidence intervals overlap 50% do not differ from chance.

Results were calculated for sixteen- and eight-note melodies separately. The hit and false alarm pairs that make up the ROC curves were derived from average participant recognition ratings for targets and lures for each participant. We used a bootstrapped significance test to compare the AUC of two ROC curves, (roc.test; Robin et al., 2011a). In pROC, bootstrapping must be used to calculate the 95% confidence intervals of the AUC when the data is obtained from “paired” ROC curves, derived from repeated measures of the same sample (Robin et al., 2011b). We used 10,000 replicates, as recommended by Carpenter and Bithell (2000) as sufficient for estimating the first significant digit (Robin et al., 2011a).

In the sixteen-note melodies, the AUC for melodies of high distinctiveness was well above chance at 81.4% (95% CI; 69.9%, 93.0%), whereas performance approximated chance for melodies of low distinctiveness at 57.0% (95% CI; 41.2%, 72.9%). A bootstrapped significance test ($n = 10,000$) revealed that the difference between the two curves was significant, $D = 2.70$, $p < .007$. Thus, a significant advantage was found in the sixteen-note melodies for distinctive over typical melodies (see Figure 4).

In the eight-note melodies, a similar pattern of improved performance for distinctive melodies was found. The AUC for high distinctiveness melodies was again above chance at 73.6% (95% CI; 59.9%, 87.3%).

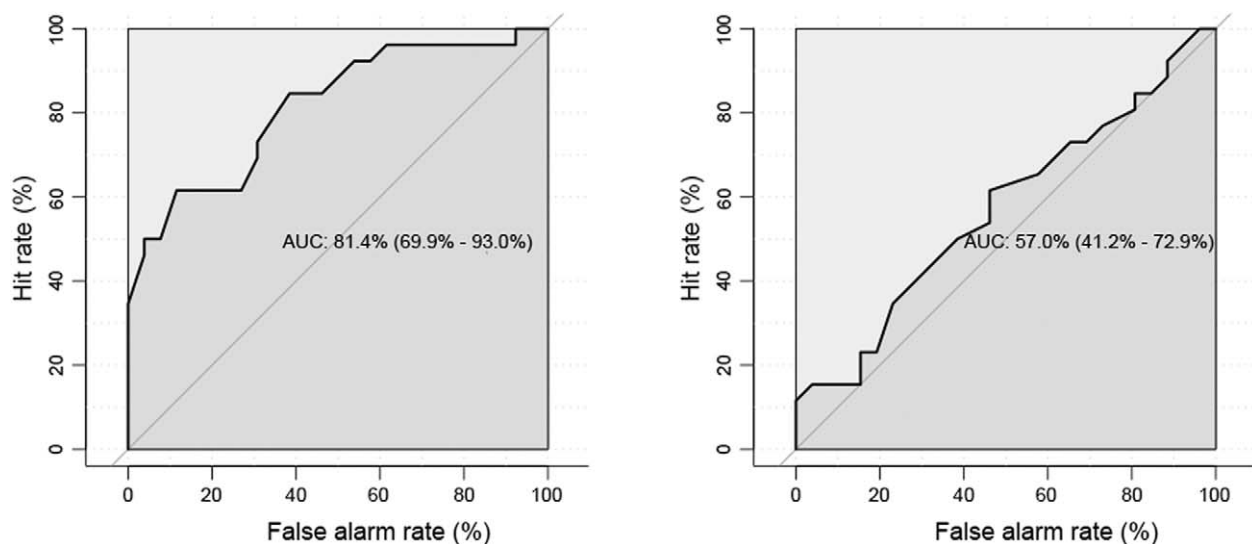


FIGURE 4. ROC curve analysis of sixteen-note melodies. A significant advantage was found for melodies of High Distinctiveness, whereas performance for Low Distinctiveness melodies was only just above chance.

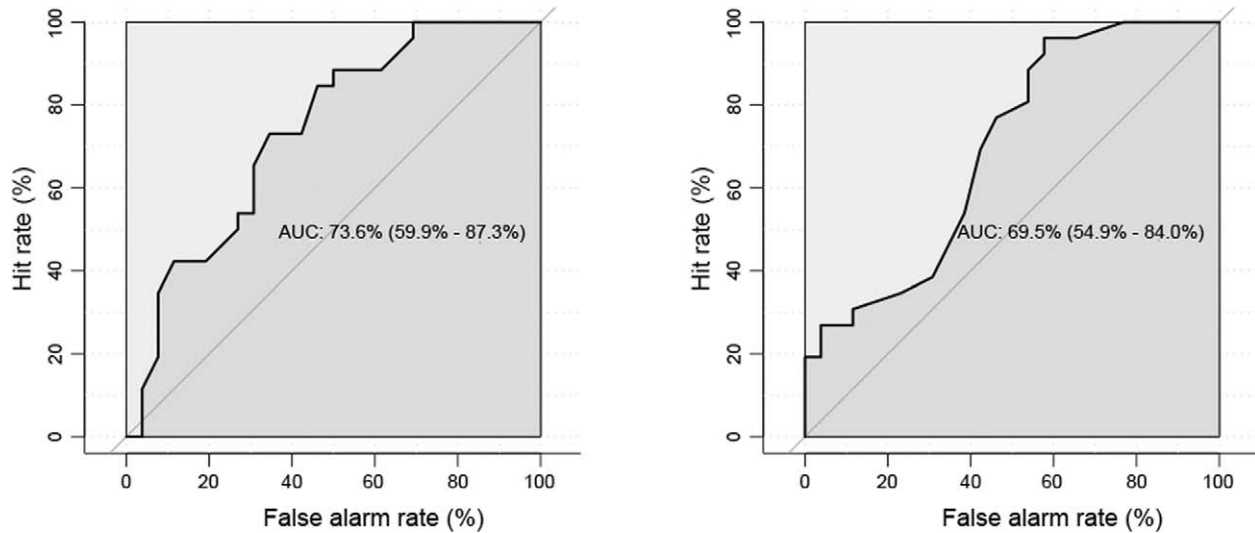


FIGURE 5. ROC curve analysis of eight-note melodies. Performance was improved in the High Distinctiveness melodies although the difference between the two groups did not reach significance.

Performance for melodies of low distinctiveness was also above chance, but with a lower AUC of 69.5% (95% CI; 54.9%, 84.0%). However, a bootstrapped significance test ($n = 10,000$) revealed that the difference between the two curves was not significant, $D = 0.51$, $p = .609$ (see Figure 5).

In summary, the ROC analyses indicate that, for sixteen-note melodies, high distinctiveness melodies were better recognized than those of low distinctiveness. Although performance was more accurate for high distinctiveness melodies of both stimulus lengths, in comparison to melodies of low distinctiveness, recognition performance did not differ between eight-note high and low distinctiveness melodies.

FACTORS PREDICTING CORRECT RECOGNITION AND FALSE MEMORIES

In the next two sections, we sought to answer two questions. First, what factors predict participants' recognition of melodies? Second, what aspects of melodic stimuli contribute to predicting recognition? Specifically, we were interested in the predictive value of melodic features outlined by the FANTASTIC framework, and whether those features that we identified in Experiment 1 as associated with the perception of distinctiveness were also associated with recognition test outcomes.

To answer these questions, we analyzed our repeated measures data using linear mixed-effects models created using the lme4 package (Bates, Maechler, Bolker, & Walker, 2013) in R (an open-source language and

environment for statistical computing; R Core Team, 2013). One of the primary benefits of this analytical approach (compared to repeated measures ANOVA) is that it allowed us to include participant and stimulus as random effects in all models (i.e., allowing random intercepts for these factors). This approach deals with the nested structure of our data (i.e., having multiple observations at each level of our manipulations within each participant), and offers important advantages in generalizing findings beyond the specific sample and stimuli tested (e.g., Baayen, Davidson, & Bates, 2008; Jaeger, 2008).

As per linear regression, chi-square tests assess whether the inclusion of a predictor significantly improves the fit of the model, and regression coefficients (b) index the degree of change in the outcome associated with a 1-unit change in the predictor (e.g., see Table 4 for analyses of recognition data). In the following analyses, we first obtained the most complex model that significantly improved fit to the data (e.g., Table 5). Coefficients from these tables are then displayed as figures (e.g., Figure 7). In these figures, only those predictors whose 95% confidence do not overlap zero are useful predictors in the best fitting version of this model.

When interpreting the coefficients in Table 4, two points are worth noting. First, we set the reference point (i.e., the intercept) as typical eight-note lures. Thus, the coefficient values in Table 4 represent the predicted increase in participants' recognition ratings for targets (cf. lures), distinctive (cf. typical) stimuli, and sixteen (cf. eight) note stimuli. Second, for linear mixed-effects

TABLE 4. Fixed Effect Coefficients for Linear Mixed-Effects Model Predicting Recognition (Experiment 2)

Fixed effect	<i>b</i>	95% CI _{<i>b</i>}	<i>t</i>
Intercept	-0.18	[-0.46, 0.08]	1.35
Status (S)	0.49*	[0.20, 0.79]	3.45
Distinctiveness (D)	0.20	[-0.13, 0.54]	1.21
Length (L)	0.19	[-0.11, 0.52]	1.14
S × D	0.08	[-0.34, 0.49]	0.38
S × L	-0.30	[-0.71, 0.07]	1.51
D × L	-0.11	[-0.58, 0.36]	0.46
S × D × L	0.56*	[0.04, 1.15]	1.98

* = 95% CIs do not include zero.

TABLE 5. Model Fit Statistics for Linear Mixed-Effects Model Predicting Recognition based on Stimulus Status and FANTASTIC Criteria (Experiment 2)

Model Step	Predictor	<i>df</i>	χ^2	<i>p</i>
1	Status (S)	1	52.58	< 0.001
2	p.range _{centred}	1	8.99	0.003
3	p.entropy _{centred}	1	0.03	0.855
4	p.std _{centred}	1	0.05	0.819
5	i.abs.range	1	3.08	0.079
6	i.abs.mean	1	0.01	0.935
7	i.abs.std _{centred}	1	0.28	0.597
8	i.mode _{centred}	1	5.98	0.014
9	i.entropy _{centred}	1	1.11	0.292
10	tonalness _{centred}	1	0.57	0.449
11	tonal.clarity _{centred}	1	1.24	0.266
12	step.cont.glob.var _{centred}	1	< 0.01	0.954
13	step.cont.loc.var _{centred}	1	3.54	0.060
14	S × p.range	1	1.26	0.262
15	S × p.entropy	1	3.24	0.072
16	S × p.std	1	0.13	0.719
17	S × i.abs.range	1	0.85	0.357
18	S × i.abs.mean	1	0.44	0.506
19	S × i.abs.std	1	0.08	0.780
20	S × i.mode	1	10.3	0.001
21	S × i.entropy	1	1.45	0.228
22	S × tonalness	1	3.68	0.055
23	S × tonal.clarity	1	< 0.01	0.978
24	S × step.cont.glob.var	1	< 0.01	0.997
25	S × step.cont.loc.var	1	0.51	0.476

models, it is not possible to generate a *p* value for the *t*-test statistic associated with each predictor. Instead, we used bootstrapping (based on 10,000 resamples of the data) to obtain 95% confidence intervals for the obtained coefficients. Confidence intervals that do not overlap zero can be taken as indicating a statistically meaningful effect.

Predicting recognition based on stimulus status, distinctiveness, and length. We created a linear mixed-effects model with recognition (i.e., participants'

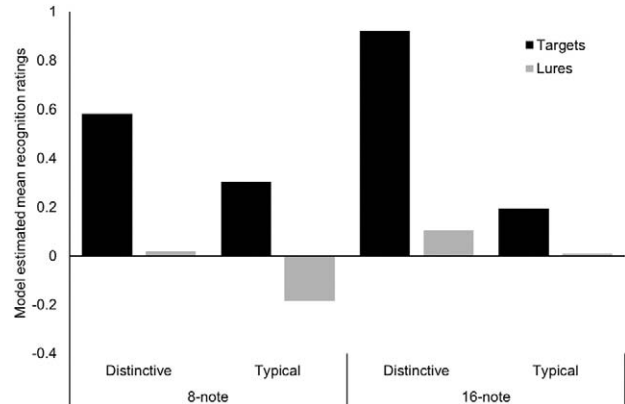


FIGURE 6. Model estimated mean recognition ratings for eight- and sixteen-note melodies when appearing as targets and lures.

recognition ratings) as the outcome, and participant and stimulus as random effects. In successive steps, we added the main effects of status (target or lure), distinctiveness, and length, and the interactions between these variables. The most complex model to improve the fit to data included all main effects and interactions, $\chi^2(1) = 3.92, p = .048$. An inspection of the coefficients in Table 4, together with the model estimated means plotted in Figure 6, indicates that targets generally received higher recognition ratings than lures (i.e., a basic memory effect), and that, consistent with the ROC curve analysis, stimulus distinctiveness only improved recognition significantly for the sixteen-note melodies.

Predicting recognition based on FANTASTIC criteria. Our analytical approach here was similar to that for the previous analysis, with one important difference. The FANTASTIC criteria were measured on continuous scales. Thus, before entering these criteria as predictors, we scaled (or standardized) them. This means the coefficients in Figure 7 indicate the change in the outcome as the value of the predictor moves away (up or down) from the mean value of the predictor, rather than as the value of the predictor moves away from zero. We have presented these coefficients in a figure, rather than using the table format used for the previous analyses, because it more clearly illustrates the predictive value of the large number of predictors included in the analysis. Panel A presents the coefficients and confidence intervals for all predictors. However, some predictors show much greater variability than others, and it was difficult to capture all predictors using a single scale. Thus, the predictors with greater variability are also presented on a rescaled figure (Panel B). Again, a specific predictor can be considered meaningful when the 95% confidence intervals do not overlap zero.

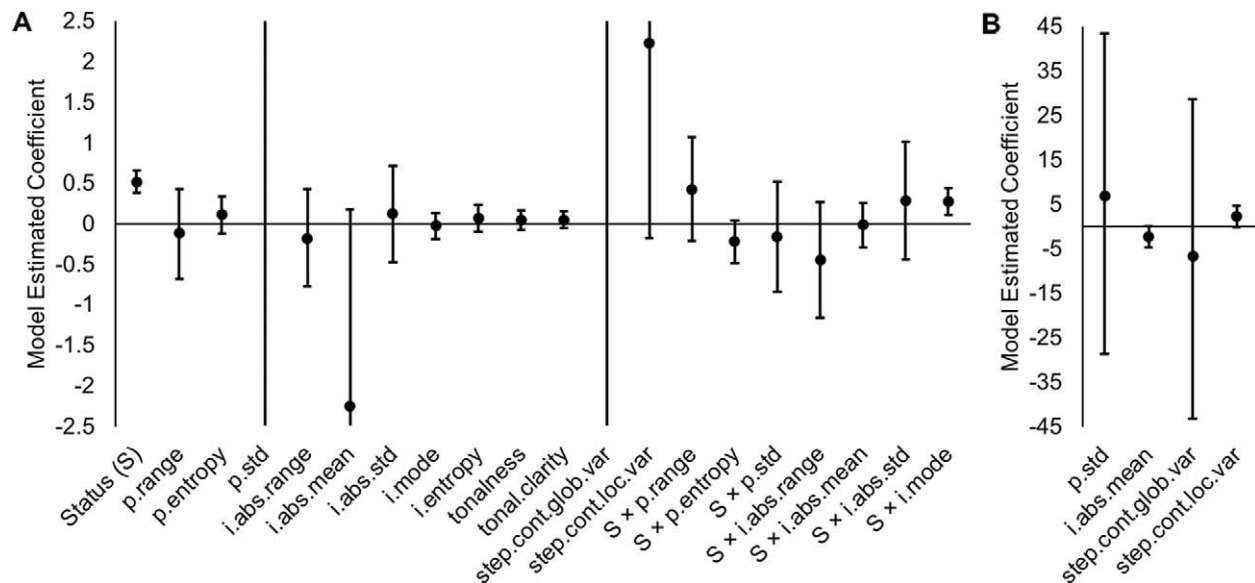


FIGURE 7. Predictors of recognition (error bars indicate 95% confidence intervals; those which do not overlap zero are useful predictors in the final version of this model). Only the interaction of stimulus status and modal interval (*i.mode*) is a significant predictor, indicating that modal interval predicted recognition in target melodies only. For clarity, Panel B shows predictors whose confidence intervals are extremely wide.

Again, we constructed a linear mixed-effects model with participants' recognition rating as the outcome variable and participant and stimulus as random effects. We then added in stimulus status (target or lure), followed by each individual FANTASTIC criterion, and the interactions between each FANTASTIC criterion and stimulus status. The most complex model to improve the fit to the data included all main effects, and a number of two-way interactions between individual FANTASTIC criteria and status (e.g., Model Step 20 in Table 5). However, as can be seen in Figure 7, only two predictors contributed significantly to recognition ratings. First, as in the previous analysis and consistent with a basic memory effect, stimulus status contributed significantly to recognition, meaning that targets received higher recognition ratings than lures. Second, the $\text{Status} \times \text{i.mode}$ interaction was significant, indicating that increases in modal interval were associated with increases in recognition ratings for targets, but not lures. Thus, an increase in the size of the most frequently used interval in a melody was associated with improved recognition (i.e., a selective increase in recognition ratings for target, but not lure, stimuli).

Features predicting hits and false alarms. We then tested whether the data provided evidence of a mirror effect, where the same features contribute to both correct recognition of targets and correct rejection of lures. To do this, we broke down the above analyses for target and lure melodies separately, to identify those FANTASTIC

criteria that predicted correct recognition of targets, and those that predicted false recognition of lures.

Following the method used above, we constructed two linear mixed-effects models. The first model used participants' recognition rating for target melodies as the outcome variable, and participant and stimulus as random effects. We then added the individual FANTASTIC criteria to the model. We repeated this procedure for the second model, using participants' recognition ratings for lure melodies as the outcome variable.

Targets. For Target melodies, as for the overall analysis, only an increase in modal interval (*i.mode*) contributed significantly to recognition ratings. Thus, a wider modal interval predicted increased hits, or correct recognition of target melodies. Conversely, a smaller modal interval predicted increased misses, or failure to recognize a target melody (see Figure 8)

Lures. We identified two factors that contributed significantly to increased false alarms, or false recognition of a lure melody as earlier heard when it was not. Interval absolute mean (*i.abs.mean*) was negatively related to false alarms, thus, as mean interval size decreased, false alarms increased. This also means that as mean interval size increased, false recognition decreased. Although this cannot be said to be a true mirror effect, because mean interval refers to the average of all intervals in the melody, rather than the most frequent interval (which contributed to correct recognition), it can be observed that, overall, an increase in the

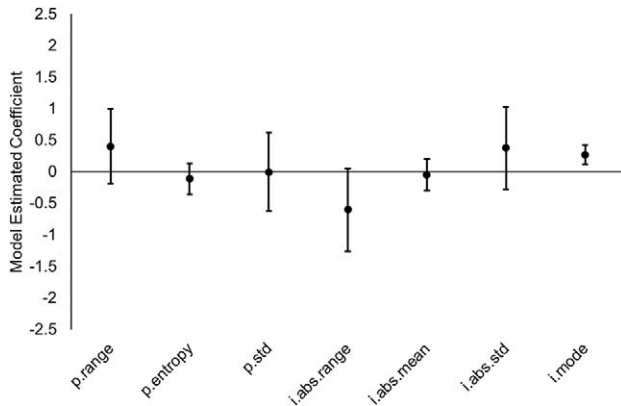


FIGURE 8. Predictors of correct recognition of target melodies (error bars indicate 95% confidence intervals). As for the overall model, modal interval (*i.mode*) predicts correct recognition of target melodies.

size of intervals used in a melody appears to contribute to both correct recognition of targets and false recognition of lure melodies.

Second, local variation in step contour (*step.cont.loc.var*) was positively related to an increase in false alarms. This means that melodies with rapid changes in contour at a local level were more likely to be falsely recognized as earlier presented (see Figure 9). Conversely, melodies with a flatter contour were less likely to generate false alarms.

In summary, while we identified a number of melodic features in Experiment 1 that contribute to the perception of a melody as distinctive, only a very small selection of these features contributed to performance on the recognition test. Wider intervals predicted both correct recognition of targets and correct rejection of lures; however, this represents only limited evidence for a mirror effect, as modal interval was associated with increased hits, whereas mean interval size across the whole melody was associated with reduced false alarms. Melodies with a rapidly changing contour were also more likely to generate false alarms, a finding that is interesting to compare with Müllensiefen and Halpern’s (2014) evidence. In their study, phrases with high repetition of unusual motifs were more likely to generate false alarms; thus, a highly varied and distinctive phrase was more likely to cause the listener to believe they had heard the melody before (Müllensiefen & Halpern, 2014). Although contour did not predict false alarms in their study, local variation in step contour is calculated from the mean absolute difference between adjacent values in the step contour vector (Müllensiefen, 2009b), and thus reflects changes in contour at the motif level. It is interesting to note that new melodies with distinctive features relating to the use of motifs were identified in both studies as more likely to be falsely identified as earlier presented.

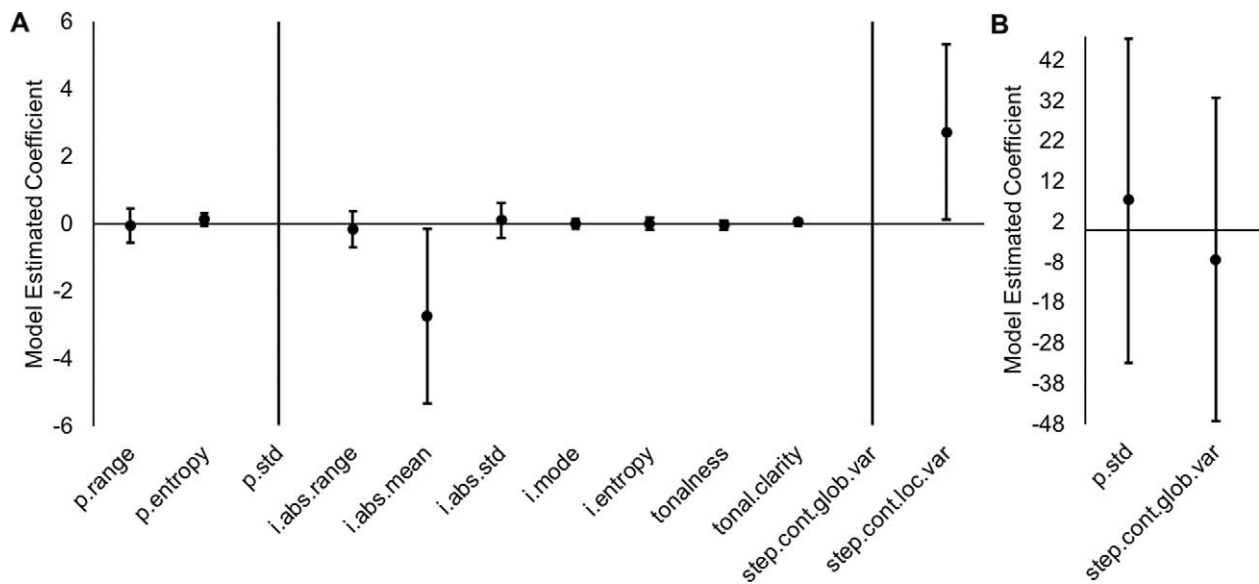


FIGURE 9. Predictors of false recognition of lure melodies (error bars indicate 95% confidence intervals). Interval absolute mean (*i.abs.mean*) is negatively related to false recognition, and local variation in step contour (*step.cont.loc.var*) is positively related to false recognition, thus, smaller intervals and varied contour predict increased false alarms. For clarity, Panel B shows predictors whose confidence intervals are extremely wide.

General Discussion

In this study, we first identified a novel stimulus set containing distinctive features, as measured through human ratings of distinctiveness. Using computational analysis, we identified a specific set of melodic features that were associated with human perception of distinctiveness in these melodies. We then used the most and least distinctive melodies in this stimulus set to test whether the distinctiveness effect, as found in recognition memory tests across many domains, could be demonstrated in music. In a recognition test of eight-note and sixteen-note melodies, the results of both ROC curve analysis and linear mixed-effects modeling confirmed that, as expected, distinctive melodies were better recognized as targets; however, this advantage was only significant in the longer (sixteen-note) melodies. Further, in examining the variables identified above as contributing to the perception of distinctiveness, only wider modal intervals predicted improved recognition performance in target melodies. To our knowledge, our study is the first to identify a specific set of musical features that are perceived as distinctive, as well as the first to demonstrate that the distinctiveness effect generalizes to recognition of whole melodies.

This result extends the findings of Bailes (2010) and Müllensiefen and Halpern (2014), who used computer-based modeling to demonstrate that improved recognition was associated with an increase in musical features that could be described as distinctive. In our study, we first used computer-based analysis as well as participant ratings to develop a set of high and low distinctiveness stimuli, and then conducted a recognition test using these stimuli to demonstrate an advantage for distinctive items in the recognition of whole melodies. Bailes (2010) identified that an increase in less-probable intervals and scale degrees (as measured using Humdrum; Huron, 1993) contributed to an earlier point-of-recognition of a known melody. This finding is consistent with our research showing that wider intervals (which would be less probable according to Bailes' measures) predict improved recognition. In addition, Bailes' (2010) findings are consistent with our findings that greater variability in pitch and intervallic content, wider intervals, and ambiguity in tonality are associated with the perception of a melody as distinctive. The less probable scale degrees identified in her study show a relationship with the more distant tones of Krumhansl and Kessler's (1982) key profiles. Tonal clarity is computed using these profiles; thus, our study shows consistency with these findings.

Although the measures that we used differed from those used by Müllensiefen and Halpern (2014), the model that they identified shows some similarities to our findings, as melodies that contained greater variety in contour, and infrequently used motifs in relation to the stimulus set were associated with improved recognition. In our study, increased variety in contour, along with increased variation in pitch and intervallic content was associated with distinctiveness. Due to the isochronous nature of our stimuli, we were unable to measure the same variables as identified in their study, and so comparison between these studies is limited. However, in music composition, the use of varying interval and pitch content would be likely to result in sequences containing unusual motifs. Further research is therefore needed using algorithms that are capable of computing first- and second-order features in isochronous as well as rhythmically complex stimuli, to build a complete model of melodic and rhythmic features that predict the perception of distinctiveness.

In this study, distinctiveness significantly predicted improved recognition only in the sixteen-note melodies. There is some precedent in the broader memory literature for a link between stimulus length and distinctiveness. For example, long words that are distinctive because they are presented in a list of short words are remembered better than short words presented in a list of long words (Hulme et al., 2006). However, in that study, items were distinctive specifically because of their length, and such results do not predict that manipulations of intrinsic distinctiveness will have stronger effects for longer stimuli than shorter stimuli. One possible explanation for our results is that the temporal nature of music allows the distinctiveness effect to accumulate over time. Bailes (2010) observed that, in addition to an advantage for momentary distinctive information, earlier points of recognition (POR) were observed where melodies contained a greater amount of distinctive material prior to the POR. Thus, it could be that longer melodies provided more scope for participants to develop a sense of distinctiveness during exposure, which may have resulted in a stronger effect of distinctiveness at test.

Some differences were found between our results and those of Vuvan and colleagues (2014). Although their study showed that highly distinctive probe tones were better recognized than moderately expectable tones, an advantage was also found in their study for highly expectable, tonally congruous probes in comparison to moderately expectable tones. Although comparison between their results and ours is limited, as our study did not incorporate a third category for moderately

distinctive melodies, one possible explanation for the differences between Vuvan and colleagues' (2014) results and ours may lie in the methodology used to investigate distinctiveness. Their study focused on investigating distinctive tonal distance in accordance with the Krumhansl and Kessler (1982) key profiles, using recognition of single probe tones, whereas our study investigated the recognition of whole melodies perceived as high or low in distinctiveness, which also showed a profile of specific melodic, intervallic, contour, and tonal features. While we found that more complex pitch, interval, and contour information contributed to the perception of distinctiveness, our investigations of implicit tonality showed a seemingly contradictory result. Although *tonal.clarity* was also negatively related to distinctiveness, meaning that tonally ambiguous melodies using less predictable and more tonally distant notes of the scale were perceived as distinctive, tonalness, or increased correlation with a Western scale, was also perceived as distinctive, even though highly tonal melodies should be highly expectable to participants with a Western musical background. It is possible that sensory priming of the modal scale used in our stimuli (Bigand et al., 2003) lead to melodies that corresponded more closely to a Western scale standing out from the corpus, and thus being rated as more distinctive by participants. Further, an advantage in recognition for harmonically distant tones was found by Vuvan and colleagues (2014) to be present only when these tones appeared within a diatonic context. This advantage disappeared when harmonically distant tones were presented within an atonal context, as these tones lost their distinctive nature. Thus, those stimuli that we composed predominantly from tones peripheral from the tonic may have been perceived as less distinctive.

A further consideration that may explain the differences between the results of our study and Vuvan and colleagues (2014) may lie in sensory versus cognitive processing of musical features. Our study showed that greater range and variability in pitch, wider intervals, and greater variability in interval size, as well as more varied contour contributed to the perception of distinctiveness. In addition, wider intervals contributed to both correct recognition of target melodies (*i.mode*) and correct rejection of lures (*i.abs.mean*). The perception of pitches and intervals is a bottom-up process that occurs at the early processing stage. Auditory information in sensory memory is grouped into events that are perceived as pitches, chords, and interval distances (Deutsch, 1999, Dowling, 1982, Snyder, 2000). Likewise, larger-scale grouping of pitches into melodic phrases gives rise to the perception of contour (Dowling, 1982;

Snyder, 2000). Tonal information, however, involves top-down, schema-based processes, and is thus a cognitive, rather than sensory, process (Krumhansl, 1991, Krumhansl & Kessler, 1982, McLachlan et al., 2013). The focus of Vuvan and colleagues' (2014) study was on the influence of tonal-schematic expectancy on memory for single tones, whereas our study combined analysis of musical features involving sensory as well as cognitive processing. The differences between their findings and ours, together with our contradictory findings regarding tonality, may therefore represent the involvement of separate sensory and cognitive processes in the perception of distinctiveness. Further research incorporating a third level of moderately distinctive information would be recommended to further understand the influence of cognitive processing of tonal information in recognition, in comparison to sensory perception of pitch, interval, and contour features.

The distinctiveness effect is normally associated with a corresponding reduction in false identifications as well as an increase in hits (Dodson & Schacter, 2001; Israel & Schacter, 1997, Schacter et al., 2001), as per the *mirror effect* (Glanzer & Adams, 1985). In this study, we found no difference between false identification of high and low distinctive lures. However, we found indirect evidence for a mirror effect in the melodic features associated with correct recognition of targets, and correct rejection of lures, as wider modal interval predicted recognition, whereas a wider mean interval predicted correct rejection. Although the variables identified are not the same, our analysis showed that overall, wider intervals are associated with both correct recognition as well as correct rejection.

In addition, we observed that greater local variation in step contour predicted an increase in false alarms. Müllensiefen and Halpern (2014) also identified that false alarms (i.e., failure to reject new melodies as lures) were associated with melodies that contain infrequently used, unusual motifs in relation to the test set. Such features are proposed to trigger false alarms because they are registered as a specially occurring event in memory. The participant might then compare this event to an existing, similar motif from another melody in the exposure phase, resulting in false recognition of the melody as earlier-heard (Müllensiefen & Halpern, 2014). It is possible that the same mechanism, where a distinctive feature in a lure melody triggers memory for a target melody, may explain distinctive contour predicting false alarms in our study. Further, as argued above, a conceptual relationship between their findings and our study cannot be ruled out, as unusual brief melodic motifs can generate increased variation in contour. Further research

is therefore needed to develop a method of measuring second-order features in isochronous melodies, which would allow us to test for this effect in our stimuli.

Although we did not find clear evidence of a mirror effect in our study, because we found no difference in false identifications of high and low distinctive lures, an absence of the mirror effect does not preclude an advantage for correct recognition of distinctive targets (Pazzaglia et al., 2014). Further, the mirror effect does not always generalize from verbal to other types of stimuli (Glanzer & Adams, 1985). As this is the first study of the distinctiveness effect in the recognition of whole melodies, further testing is required to determine whether the small sample size contributed to a lack of an effect of distinctiveness on false alarms. However, Müllensiefen and Halpern (2014) also did not find evidence of a mirror effect for distinctive items, as some differences were identified in the factors that predicted correct rejection of lures, in comparison to those that contributed to recognition of old items. Although infrequently used (i.e., distinctive) motifs were identified as a factor in both models, contrary to the mirror effect, these were more likely to elicit judgements of a melody as previously heard in both targets and lures. This correlation between features predicting both hits and false alarms was also found by Cortese, Khanna, and Hacker (2010) when studying the effect of word frequency on recognition, suggesting that sublexical processes (e.g., orthographic and phonological processing) may gain importance when semantic processing is not possible, for example when remembering nonwords. Müllensiefen and Halpern (2014) proposed that, because unfamiliar nonverbal music lacks a semantic dimension, sublexical processes may also be associated with memory for nonverbal music. Our study also used nonverbal melodies; however, hits and false alarms did not correlate, although we did identify a separate set of features contributing to false alarms. We were unable to test for the same features measured in their study due to the isochronous rhythm of our stimuli. Nevertheless, together with Müllensiefen and Halpern's (2014) study, our results contribute to an emerging finding that the distinctiveness effect operates differently in musical recognition to other stimulus types.

Several limitations must be identified in this study. The MUSOS Toolkit (Rainsford et al., 2018) uses a step-sequencer to visually display melodic contours to the participant during exposure. Although we removed the visual display of melodies during the test phase, it is possible that the use of a step-sequencer interface during encoding might cause participants to perceive a target

melody as distinctive based on its visual features. If this was to occur, two potential outcomes are possible. First, that the high distinctiveness melodies were rich in both distinctive visual and auditory features in comparison to low distinctiveness melodies, thus the visual display might exaggerate the effect of distinctiveness on melodic recognition. A second possible outcome might be a less-clear differentiation between high and low distinctive melodies, as melodies low in distinctive musical features might contain separate visually distinctive features. Thus, further testing without the step-sequencer interface would be beneficial to clarify the effect of distinctiveness on melodic recognition, in particular for eight-note melodies, where the result was not significant.

Vuvan and colleagues (2014) found that both highly expectable, as well as highly unexpected tones in relation to the diatonic scale were better remembered in comparison to moderately expectable tones. However, our stimulus set contained only melodies that were high and low in distinctive content; thus, the absence of a category for moderately distinctive melodies in our study limits comparison with their findings. If both highly expectable and highly distinctive features contributed to melodic recognition, arguably, our study would show no difference between the high and low distinctiveness melody collections, whereas we found an advantage for longer melodies with distinctive content. Vuvan and colleagues (2014) did, however, only interpret improved recognition of highly unexpected tones as consistent with a distinctiveness heuristic. Improved memory for highly expectable tones was instead explained in connection with an availability heuristic; thus, their study identified two separate mechanisms that contribute to improved musical recognition. Further research is therefore needed to identify whether the musical features which are associated with perception of distinctiveness and expectability differ, and the role of these separate mechanisms in predicting recognition performance.

This study represents a further contribution towards the understanding of the role of distinctiveness in the recognition of musical material. We identified a specific set of musical features that were associated with the perception of distinctiveness in brief melodies. These features show relationships with those identified in recent research as contributing to improved melodic recognition, such as infrequently used pitches and intervals (Bailes, 2010) and variety in contour (Müllensiefen & Halpern, 2014). Further, our findings suggest that, as for other domains, whole melodies that are rich in distinctive features are better recognized, and identified with greater accuracy, than those of low distinctiveness. However, the number of studies examining

distinctiveness in music is still very low. Those that have found an advantage for distinctive items have used different computer-based techniques to identify distinctive characteristics of musical material. Although the factors that these studies have identified are similar, further research is needed to develop a more complete model of melodic features that are perceived as distinctive. Further studies replicating the distinctiveness effect, and investigating factors that lead to false alarms as well as correct recognition, are also needed in order for the distinctiveness effect in music to be fully understood.

Author Note

This research was funded by Australian Research Council Discovery Grant DP140103746 to M. A. Palmer et al. and an Australian Government Research Training Program Scholarship to M. Rainsford.

Correspondence concerning this article should be addressed to Miriam Rainsford, Psychology, School of Medicine, University of Tasmania, Private Bag 30, Hobart, Tasmania, 7001, Australia. E-mail: miriam.rainsford@utas.edu.au

References

- BAAYEN, R. H., DAVIDSON, D. J., & BATES, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. DOI: 10.1016/j.jml.2007.12.005
- BAILES, F. (2010). Dynamic melody recognition: Distinctiveness and the role of musical expertise. *Memory and Cognition*, 38, 641–650. DOI: 10.3758/MC.38.5.641
- BATES, D. M., MAECHLER, M., BOLKER, B., & WALKER, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4 classes*. Retrieved from <http://lme4.r-forge.r-project.org>
- BIGAND, E., POULIN, B., TILLMANN, B., MADURELL, F., & D'ADAMO, D. A. (2003). Sensory versus cognitive components in harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 159–171. DOI: 10.1037.0096-1523.29.1.159
- BRANDT, K. R., GARDINER, J. M., & MACRAE, C. N. (2006). The distinctiveness effect in forenames: The role of subjective experiences and recognition memory. *British Journal of Psychology*, 97, 269–280. DOI: 10.1348/000712605X73685
- BÜLTHOFF, I., & NEWELL, F. N. (2015). Distinctive voices enhance the visual recognition of unfamiliar faces. *Cognition*, 137, 9–21. DOI: 10.1016/j.cognition.2014.12.006
- CARPENTER, J., & BITHELL, J. (2000). Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, 19, 1141–1164. DOI: 10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F
- COHEN, M. E., & CARR, W. J. (1975). Facial recognition and the von Restorff effect. *Bulletin of the Psychonomic Society*, 6, 383–384. DOI: 10.3758/BF03333209
- CORTESE, M. J., KHANNA, M. M., & HACKER, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory*, 18, 595–605. DOI: 10.1080/09658211.2010.493892
- CURRAN-EVERETT, D. (2000). Multiple comparisons: Philosophies and illustrations. *American Journal of Physiology: Regulatory, Integrative, and Comparative Physiology*, 279, R1–R8.
- CYCLING '74 (2014). *Max/MSP 6.1* [Computer software]. Walnut, CA: Cycling '74. Retrieved from <http://www.cycling74.com>
- DIENES, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, 6, 274–290. DOI: 10.1177/1745691611406920
- DEUTSCH, D. (1999). The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 349–411). New York: Academic Press.
- DEWHURST, S. A., & PARRY, L. A. (2010). Emotionality, distinctiveness, and recollective experience. *European Journal of Cognitive Psychology*, 12, 541–551. DOI: 10.1080/095414400750050222
- DODSON, C. S., & SCHACTER, D. L. (2001). “If I had said it I would have remembered it”: Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin and Review*, 8, 155–161. DOI: 10.3758/BF03196152
- DOWLING, W. J. (1982). Melodic information processing and its development. In D. Deutsch (Ed.), *The psychology of music* (1st ed., pp. 413–429). New York: Academic Press.
- GLANZER, M., & ADAMS, J. K. (1985). The mirror effect in recognition memory. *Memory and Cognition*, 13, 8–20. DOI: 10.3758/BF03198438
- HÉBERT, S., & PERETZ, I. (1997). Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? *Memory and Cognition*, 25, 518–533. DOI: 10.3758/BF03201127
- HULME, C., NEATH, I., STUART, G., SHOSTAK, L., SURPRENANT, A. M., & BROWN, G. A. (2006). The distinctiveness of the word-length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 586–594. DOI: 10.1037/0278-7393.32.3.586
- HURON, D. (1993). *The Humdrum Toolkit: Software for Music Researchers*. [Computer software]. Stanford, CA: Center for Computer Assisted Research in the Humanities
- ISRAEL, L., & SCHACTER, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin and Review*, 4, 577–581. DOI: 10.3758/BF03214352

- JAEGER, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434-446. DOI: 10.1016/j.jml.2007.11.007
- JAKUBOWSKI, K., FINKEL, S., STEWART, L., & MÜLLENSIEFEN, D. (2016). Dissecting an earworm: Melodic features and song popularity predict involuntary musical imagery. *Psychology of Aesthetics, Creativity, and the Arts*, 11, 122-135. DOI: 10.1037/aca0000090
- JEFFREYS, H. (1961). *The theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.
- JOHNSON-LAIRD, P. N., KANG, O. E., & LEONG, Y. C. (2012). On musical dissonance. *Music Perception*, 30, 19-35. DOI: 10.1525/mp.2012.30.1.19
- KAUSLER, D. H., & PAVUR, E. J. (1974). Orthographic distinctiveness of consonants and recognition learning. *Journal of Experimental Psychology*, 102, 435-438. DOI: 10.1037/h0035893
- KRUMHANSL, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford, UK: Oxford University Press.
- KRUMHANSL, C. L. (1991). Music psychology: Tonal structures in perception and memory. *Annual Review of Psychology*, 42, 277-303. DOI: 10.1146/annurev.ps.42.020191.001425
- KRUMHANSL, C. L., & KESSLER, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89, 334-368.
- MCLACHLAN, N., MARCO, D., LIGHT, M., & WILSON, S. (2013). Consonance and pitch. *Journal of Experimental Psychology: General*, 142, 1142-1158. DOI: 10.1037/a0030830
- MICKES, L., FLOWE, H. D., & WIXTED, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*, 18, 361-376. DOI: 10.1037/a0030609
- MÜLLENSIEFEN, D. (2009a). *FANTASTIC: Feature ANalysis Technology Accessing SStatistics (In a Corpus)*. [Computer software]. London, UK: Goldsmiths, University of London. Retrieved from <http://www.doc.gold.ac.uk/isms/m4s/FANTASTIC.zip>
- MÜLLENSIEFEN, D. (2009b). *FANTASTIC: Feature ANalysis Technology Accessing SStatistics (In a Corpus): Technical Report v1.5*. London, UK: Goldsmiths, University of London. Retrieved from http://www.doc.gold.ac.uk/isms/m4s/FANTASTIC_docs.pdf
- MÜLLENSIEFEN, D., & HALPERN, A. (2014). The role of features and context in recognition of novel melodies. *Music Perception*, 31, 418-455. DOI: 10.1525/MP.2014.31.5.418
- PAZZAGLIA, A. M., STAUB, A., & ROTELLO, C. M. (2014). Encoding time and the mirror effect in recognition memory: Evidence from eyetracking. *Journal of Memory and Language*, 75, 77-92. DOI: 10.1016/j.jml.2014.05.009
- R CORE TEAM. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- RAINSFORD, M., PALMER, M. A., & PAINE, G. (2018). The MUSOS (MUSIC Software System) Toolkit: A computer-based, open source application for testing memory for melodies. *Behavior Research Methods*, 50(2), 284-702. DOI: 10.3758/s13428-017-0894-6
- RAJARAM, S. (1998). The effects of conceptual salience and perceptual distinctiveness on conscious recollection. *Psychonomic Bulletin and Review*, 5, 71-78. DOI: 10.3758/BF03209458
- ROBIN, X., TURCK, N., HAINARD, A., TIBERTI, N., LISACEK, F., SANCHEZ, J.-C., & MÜLLER, M. (2011a). *Package 'pROC'*. Retrieved August 31, 2017, from http://web.expasy.org/pROC/files/pROC_1.7.2_R_manual.pdf
- ROBIN, X., TURCK, N., HAINARD, A., TIBERTI, N., LISACEK, F., SANCHEZ, J.-C., & MÜLLER, M. (2011b). pROC: An open-source package for R and S+ to analyse and compare ROC curves. *BMC Bioinformatics*, 12, 77. DOI: 10.1186/1471-2105-12-77.
- ROUDER, J. N., SPECKMAN, P. L., SUN, D., MOREY, R. D., & IVERSON, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, 16, 225-237. DOI: 10.3758/PBR.16.2.225
- SCHACTER, D. L., CENDAN, D. L., DODSON, C. S., & CLIFFORD, E. R. (2001). Retrieval conditions and false recognition: Testing the distinctiveness heuristic. *Psychonomic Bulletin and Review*, 8, 827-833. DOI: 10.3758/BF03196224
- SCHACTER, D. L., & WISEMAN, A. L. (2006). Reducing memory errors: The distinctiveness heuristic. In R. R. Hunt (Ed.), *Distinctiveness and memory* (pp. 89-107). New York: Oxford University Press.
- SCHMUCKLER, M. A. (1997). Expectancy effects in memory for melodies. *Canadian Journal of Experimental Psychology*, 51, 292-306. DOI: 10.1037/1196-1961.51.4.292
- SHANNON, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, 27, 379-423, 623-656. Updated version available at <http://math.harvard.edu/ctm/home/text/others/shannon/entropy/entropy.pdf>
- SLOBODA, J. A., & PARKER, D. H. (1985). Immediate recall of melodies. In P. I. C. Howell, & R. West (Eds.), *Musical structure and cognition* (pp. 143-167). London, UK: Academic Press.
- SNYDER, B. (2000). *Music and memory: An introduction*. Cambridge, MA: MIT Press.
- STEVENS, C., & BYRON, T. (2009). Universals in music processing. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 14-23). Oxford, UK: Oxford University Press.
- SWETS, J. A. (1973). The relative operating characteristic in psychology. *Science*, 182, 990-1000.
- TEMPERLEY, D. (2007). *Music and probability*. Cambridge, MA: MIT Press.

VALENTINE, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, 43, 161–204. DOI: 10.1080/14640749108400966

VUVAN, D. T., PODOLAK, O. M., & SCHMUCKLER, M. A. (2014). Memory for musical tones: The impact of tonality and the creation of false memories. *Frontiers in Psychology*, 5, 1–18. DOI: 10.3389/fpsyg.2014.00582

VON RESTORFFE, H. (1933). Über die virkung von bereichsbildungen im spurenfeld [The effects of field formation in the trace field]. *Psychologische Forschung*, 18, 299–342.

WAGENMAKERS, E.-J., WETZELS, R., BORSBOOM, D., VAN DER MAAS, H. L. J., & KIEVIT, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7, 632–638. DOI: 10.1177/1745691612463078