Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23-28, 2012, Thessaloniki, Greece
Cambouropoulos E., Tsougras C., Mavromatis P., Pastiadis K. (Editors)

A Melodic Similarity Measure Based on Human Similarity Judgments

Naresh N. Vempala,*[1] Frank A. Russo*[2]

*Department of Psychology, Ryerson University, Canada

[1]nvempala@psych.ryerson.ca, [2]russo@psych.ryerson.ca

ABSTRACT

Music software applications often require similarity-finding methods. One instance involves performing content-based searches, where music similar to what is heard by the listener is retrieved from a database using audio or symbolic input. Another instance involves music generation tools where compositional suggestions are provided by the application based on user-provided musical choices (e.g. genre, rhythm and so on) or samples. The application would then generate new samples of music with varying degrees of musical similarity. Although several similarity algorithms such as edit distance methods and hidden Markov models already exist, they are not fully informed by human judgments. Furthermore, only a few studies have compared human similarity judgments with algorithmic judgments. In this study, we describe an empirically derived measure, from participant judgments based on multiple linear regression, for determining similarity between two melodies with a one-note change. Eight standard melodies of equal duration (eight notes) were systematically varied with respect to *pitch distance, pitch direction, tonal stability, rhythmic salience,* and *melodic contour*. Twelve comparison melodies with one-note changes were created for each standard. These comparison melodies were presented to participants in transposed and non-transposed conditions. For the non-transposed condition, predictors of similarity were pitch distance, direction and melodic contour. For the transposed condition, predictors were tonal stability and melodic contour. In a follow-up experiment, we show that our empirically derived measure of melodic similarity yielded superior performance to the Mongeau and Sankoff similarity algorithm. We intend to extend this measure to comparison melodies with multiple note changes.

I. INTRODUCTION

Music software applications require the use of efficient similarity-finding methods for various reasons. Two important instances when similarity finding methods are required are (a) in content-based music search engines, where pieces of music or songs similar to the one being heard by the listener need to be retrieved from a database based on symbolic or audio input (Clausen & Kurth, 2002; Pauws, 2002; Prechelt & Typke, 2001; Typke, Wiering, and Veltkamp, 2005), and (b) in music composition tools that provide compositional suggestions based on user-provided choices such as genre, rhythm and so on. For example, a music generation application might take a specific sample of music as input and generate new samples with varying degrees of melodic and rhythmic similarity. Ideally, the application would generate a ranked list of selections based on similarity with the source sample.

Many similarity-finding algorithms already exist in the domain of music information retrieval. Some of these algorithms are string matching algorithms such as edit distance (McNab, Smith, Witten, Henderson, and Cunningham, 1996; Wagner & Fischer, 1974; Mongeau & Sankoff, 1990), while some are probabilistic such as Markov and, and hidden Markov models (Kim, Lee, Yoon, and Lee, 2008). Although a few previous studies have compared human similarity judgments with algorithmic judgments (Eerola, Järvinen, Louhivuori, and Toiviainen, 2001; Müllensiefen & Frieler, 2004a, 2004b, 2007), applications in the domain of music information retrieval that use similarity measures are still fairly devoid of cognitive approaches informed by human judgments. In the extensive comparison study conducted by Müllensiefen and Frieler (2004a), melodic similarity measurements computed by 50 algorithms were compared with similarity judgments of 99 human participants. The stimuli consisted of 14 melodies and 84 variants of these melodies. Müllensiefen and Frieler found that edit distance measurements with a rich symbolic representation compared well to human similarity judgments.

Our goal in this study was to extend beyond the mere comparison of empirical and algorithmic judgments of musical similarity, and to provide an empirically derived measure for determining similarity between two melodies based on human judgments. A method based purely on human judgment could either be implemented within an existing music analysis/generation application as is, or added as a higher level similarity measure, on top of an existing set of low level similarity algorithms, prioritized under certain conditions. This approach would allow the application to resemble human cognition when evaluating for similarity.

In this paper, we describe our method for determining similarity based on five separate musical predictors. We consider these five predictors as quasi-independent variables that affect human similarity judgments with respect to melodies. While several low level acoustic features within the audio signal may play a role in affecting human judgments of similarity across different pieces of music, we restricted our set of predictors to five high level musical concepts: *pitch distance*, *pitch direction*, *rhythmic salience*, *melodic contour*, and *tonal stability*. These five predictors were chosen because of their importance in capturing the musical content within a melody as perceived by a listener.

Empirical studies have demonstrated the importance of rhythm as well as pitch when linking musical structure to cognitive processes (Krumhansl, 2000). Smith and Cuddy (1989) found that pitch changes were easier to detect in the context of a binary meter as compared to a ternary meter. Prince, Schmuckler, and Thompson (2009) found that participants were more accurate at making timing judgments when a tonally stable probe occurred on a metrically stable position or when a tonally unstable probe occurred on a metrically unstable position. A number of studies by Dowling and colleagues (1971, 1972, 1978) have demonstrated the importance of contour in melody recognition, particularly in the case of transposed melodies. Schubert and Stevens (2006) showed that listeners use pitch distance as an important factor for determining similarity across melodies. Based on evidence from these and related studies, we selected our set of five predictors. Our method for deriving a similarity-finding measure was based on multiple linear regression as represented by the following equation.

$$y = mx_1 + mx_2 + mx_3 + mx_4 + mx_5 + B \quad (1)$$

Here, y is the magnitude of similarity on a scale from 1 (least similar) to 5 (most similar); x_1, x_2, x_3, x_4, and x_5 are the values of the five predictors; m is the slope corresponding to the beta weight for each predictor; and B is the intercept (constant).

II. METHOD AND MATERIALS

Twenty-two participants (15 female) from the Ryerson community participated in this experiment. Participants received credit for an introductory psychology class. They had 0 to 15 years of music training (M = 4.3, SD = 4.4) and ranged in age from 18 to 41 years (M = 23.3, SD = 8.3). Each participant heard an eight-note *standard* melody followed by an eight-note variation of the standard. The variation was referred to as the *comparison* melody. After listening to the comparison melody, listeners rated the level of similarity between the standard and comparison melodies on a Likert-type scale from 1 (least similar) to 5 (most similar).

We composed eight standard melodies. Four of these melodies were in the C major scale, while the remaining four were in the C minor scale. All melodies were of equal duration and consisted of eight isochronous notes. To ensure that the eight standard melodies were characteristic of real melodies with tonal organization, we verified them by correlation with the Krumhansl and Kessler profiles (Krumhansl & Kessler, 1982), and resolution to the tonic.

We created 12 unique comparison melodies (variants) for each standard by systematically manipulating the standard with respect to rhythmic salience, pitch distance, and pitch direction. Three of the 12 variants were created by increasing the pitch of the fourth note in the standard melody by one, two, or three scale notes. Likewise, three more variants were created by decreasing the pitch of the fourth note in the standard melody by one, two, or three scale notes. Six additional variants were created by adopting the same procedure on the fifth note of the standard melody instead of the fourth note.

Each stimulus consisted of three bars. A percussive rhythm track was played for the entire duration of all three bars. The standard melody was played in the first bar, followed by a bar of rest, followed by a bar of the comparison melody. The rhythm track that accompanied the standard and comparison melodies consisted of eight isochronous eighth note clicks per bar representing an 8-note pulse at a tempo of 120 bpm. Clicks 1, 3, 5, and 7 were strongly accented, while clicks 2, 4, 6, and 8 were weakly accented. This alternating pattern of strong and weak accents provided a clear binary metric framework. Since each melody was eight notes in duration, each eighth note click corresponded to a note-onset in the melody. The design of the comparison melodies as described above enabled us to vary the levels of the five predictors in the following manner.

(1) *Rhythmic salience*: On the basis of short-term memory theory (Murdock Jr., 1962), we chose to make pitch manipulations only on notes 4 and 5 as opposed to notes closer to the beginning or end of the melodic sequence. Choosing notes towards the middle of the melodic sequence allowed us to avoid any confounding effects associated with primacy or recency. Since note 4 in the melody occurred on a weakly accented position and note 5 occurred on a strongly accented position, a change to note 4 in the comparison melody was regarded as a change occurring on a position of weak rhythmic salience (coded as 0), while a change to note 5 was regarded as a change occurring on a position of strong rhythmic salience (coded as 1).

(2) *Pitch distance*: Degree of change in pitch with respect to the standard melody was coded as 1, 2, or 3 corresponding to a change by one, two, or three scale tones respectively.

(3) *Pitch direction*: Direction of pitch change with respect to the standard melody was captured by coding an increase as 1 and a decrease as -1.

(4) *Tonal stability*: Since all melodies were either in C major or in C minor, notes in the major and minor scales were divided into three levels of tonal stability. The tonic, C, was the most stable. The third and the fifth scale notes (E, G for major; Eb, G for minor) fell into the second level of stability. All remaining notes in the scale fell into the third level, which was the least stable. The changed note in the comparison melody was coded for change in tonal stability with respect to the standard. Codes ranged from -2 to 2, depending on the magnitude and direction of the change. No change in tonal stability was coded as 0.

(5) *Melodic contour*: Change in melodic contour was coded with respect to the changed note (4 or 5), the note preceding it, and the note succeeding it. For example, if note 4 was changed in the comparison melody, then contour change was examined by taking notes 3, 4, and 5 into consideration. Contour changes were coded based on Narmour's implication-realization model (1990). A change in contour was coded as 1, and no change was coded as 0.

Figure 1 provides examples of three different cases illustrating how contour changes were coded. In the first melodic segment (a), note 4 decreases in pitch in the comparison melody. Hence, pitch direction between notes 3-4-5 changes from *up-down* in the standard to *same-same* in the comparison and is considered a change in melodic contour. In the second melodic segment (b), note 4 increases in pitch in the comparison melody. Hence, pitch direction between notes 3-4-5 changes from *up-up* in the standard to *up-same* in the comparison. On the basis of the implication-realization model, this is not considered as a change in melodic contour. In the third melodic segment (c), note 4 decreases in pitch. Hence, pitch direction between notes 3-4-5 changes from *same-down* to *down-down* in the comparison. Taking into account the ascending pitch direction from 2-3, 3-4 would be interpreted in the standard melody as going *up*. The change in pitch direction occurring between notes 3-4-5 of the standard (*up-down*) ceases to be a change in pitch direction in the comparison (*down-down*) and is therefore considered as a change in melodic contour.

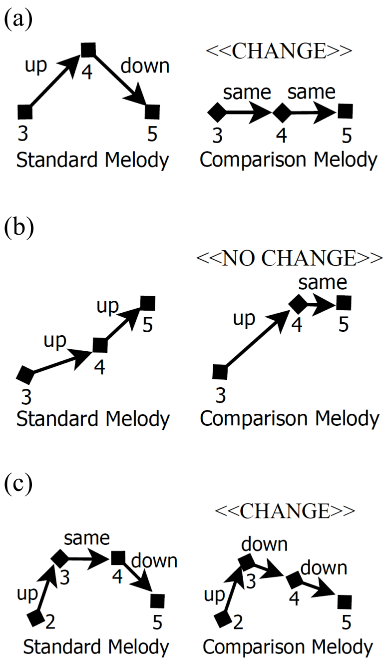


Figure 1. (a) Pitch decrease in note 4 and change in melodic contour. (b) Pitch increase in note 4 and no change in melodic contour. (c) Pitch decrease in note 4 and change in melodic contour.

Examples of standard melodies and their variants, and the coding of predictor variables for each of these variants are provided in Table 1. In addition to making similarity judgments of comparison melodies in the same key as the standard melody, we also asked participants to make similarity judgments for comparison melodies that were transposed. When a melodic fragment is repeated under transposition, listeners are generally able to appreciate that the fragment has been repeated despite its presentation using a novel set of pitch classes. In certain contexts, while the main parts of a composition (such as the chorus and bridge) retain the same key, a motif may be repeated in a transposed key generating a sense of perceptual novelty while still retaining a sense of similarity (Van Egmond, Povel, and Maris, 1996). Hence, our motivation for presenting transposed melodies to participants in addition to same-key melodies was to understand how the five predictors affect similarity judgments across transpositions, and how their effect is different from non-transposed comparisons.

Prior to listening to the stimuli, participants went through a training phase where they familiarized themselves with the eight standard melodies. In the training phase, participants were made to listen to each of the eight melodies for six repetitions. After the training phase, participants heard all the stimuli in two blocks, one consisting of all the major melodies and one consisting of all the minor melodies. Each block contained 96 stimuli. 48 stimuli consisted of same-key comparisons (12 variants x 4 standard melodies). The remaining 48 stimuli were transposed versions of the same variants. All 96 stimuli were presented in a randomized order such that participants were unaware prior to listening as to whether the comparison melody would be a same-key or a transposed variant.

Table 1. Example of one variant for each of two standard melodies, with coded values for the five predictor variables. Here *SM* = Standard Melody, *VT* = Variant Type, *CM* = Comparison Melody, *D* = Pitch Distance, *Di* = Pitch Direction, *T* = Tonal Stability, *R* = Rhythmic Salience, *C* = Melodic Contour.

SM	VT	CM	D	Di	T	R	C
G F G A F E D C	Note 4 decreased by 1 scale note	G F G **G** F E D C	1	-1	1	0	0
G F G A F E D C	Note 5 decreased by 3 scale notes	G F G A **C** E D C	3	-1	2	1	1

The procedure used for creating transposed versions and presenting them to participants is described as follows. Previous work using transposed melodies suggests that when a melody is transposed, as pitch distance increases and the number of shared pitches decreases from the current key to the transposed key, the level of difficulty in perceiving the similarity between the two melodies increases (Frances, 1958, as cited in Dowling, 1978). Following previous work by Dowling (1978), we transposed all the melodies up by a major third (C to E) and down by a minor third (C to A). Four different transposition sets were created where for each of the eight standard melodies, the comparison melodies were either transposed up or transposed down as described here. In Set 1, all comparisons from standard melodies 1 and 4 in C major and C minor were transposed up, while comparisons from standard melodies 2 and 3 in C major and C minor were transposed down. In Set 2, all comparisons from standard melodies 2 and 3 in C major and C minor were transposed up, while comparisons from standard melodies 1 and 4 in C major and C minor were transposed down. In Set 3, all comparisons from standard melodies 1 and 2 in C major and C minor were transposed up, while comparisons from standard melodies 3 and 4 in C major and C minor were transposed down. In Set 4, all comparisons from standard melodies 3 and 4 in C major and C minor were transposed up, while comparisons from standard melodies 1 and 2 in C major and C minor were transposed down. Participants were assigned to any one of the four sets.

III. RESULTS AND DISCUSSION

A mixed design analysis of variance (ANOVA) was performed for major and minor melodies using within-subjects factors of melody (1-4) and variant (1-12), and a between-subjects factor of transposition set. As expected, our results indicated no between-subjects effect due to transposition set. Separate linear regressions were then performed on the non-transposed and transposed melody comparisons. For non-transposed melodies, we performed stepwise regression using the five predictors as independent variables and the mean similarity ratings for each specific type of comparison melody as the dependent variable. This method allowed us to examine which of the predictor variables were strongly correlated with the similarity ratings, and include only the strongly correlated variables in the regression model. The model that emerged from the stepwise analysis contained only three predictor variables: pitch distance, pitch direction, and melodic contour. The model was significant (*F(1,91) = 11.3, p < .001*) and accounted for 26.9 % of the variance in the similarity values. Pitch distance and pitch direction (*p < .01, p < .01* respectively) were significant as predictors in the model. Melodic contour was marginally significant (*p = .06*). When tonal stability was entered into the model, the increase in explained variance was negligible, tonal stability was not significant as a predictor (*p = .25*), and melodic contour became a significant predictor (*p < .05*). Thus we arrived at our formula for measuring similarity between a standard and a comparison melody with a one-note change by applying the values obtained in the initial regression involving stepwise entry of the variables:

$$Y_{NONTRANSPOSED} = -0.12x_1 + 0.17x_2 - 0.13x_3 + 3.56 \quad (2)$$

Here $Y_{NONTRANSPOSED}$ is the similarity value, x_1 is the change in pitch distance of the note between the standard and comparison melodies, x_2 is the direction of its pitch change, and x_3 is the change in melodic contour. The contribution of each predictor, to the overall variance, is shown in Figure 2.

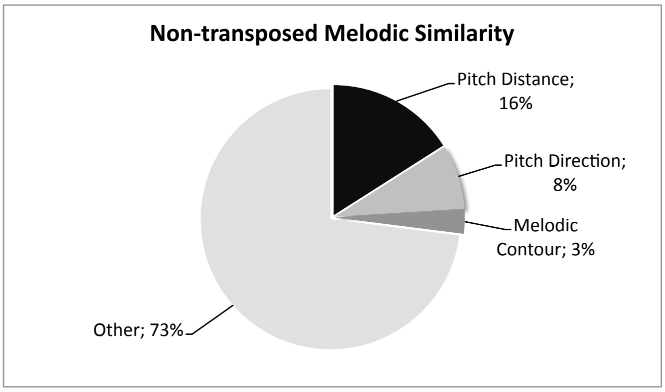


Figure 2. Percentage of variance contribution of each predictor variable in the similarity-finding model for non-transposed melodies.

For transposed melodies, we performed stepwise regression similar to the non-transposed melodies using the five predictors as independent variables and the mean similarity ratings for each specific type of comparison melody as the dependent variable. The model that emerged from the stepwise regression analysis contained only two predictor variables: tonal strength and melodic contour. The model was significant (*F(1,92) = 12.9, p < .001*) and accounted for 21.7 % of the variance in the similarity values. The inclusion of pitch distance and pitch direction within the model did not result in any noticeable improvement. However, although rhythmic salience was not a significant predictor (*p = .15*), the model's performance improved slightly with its inclusion. With these three predictors the model accounted for 23.4% of the variance in the similarity values. Since this increase was still low, we opted for the two-predictor model consisting of tonal strength and melodic contour, which were both significant as predictors (*p < .05, p < .001* respectively). The formula for measuring similarity between a standard and a transposed comparison melody with a one-note change was thus determined by applying the values obtained in the initial regression involving stepwise entry of the variables:

$$Y_{TRANSPOSED} = -0.057x_1 - 0.27x_2 + 2.85 \quad (3)$$

Here $Y_{TRANSPOSED}$ is the similarity value, x_1 is the change in tonal strength of the note between the standard and comparison melodies, and x_2 is the change in melodic contour. The contribution of each predictor, to the overall variance, is shown in Figure 3.

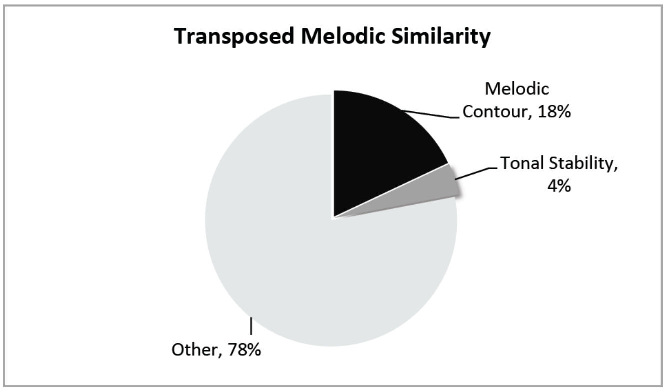


Figure 3. Percentage of variance contribution of each predictor variable in the similarity-finding model for transposed melodies.

Prior to running the regression analysis, we assumed that both rhythmic salience and tonal stability would have an effect on the similarity judgments of participants for non-transposed melodies. Specifically, with respect to rhythmic salience we expected that if a note was altered on a strong beat (note 5) instead of a weak beat (note 4), the change on the strong beat would be interpreted by the listener as more salient. Hence, changes in similarity between the standard and the comparison melodies would be perceived more easily when occurring on a strong beat.

There are three possible reasons why rhythmic salience and tonal salience were not reflected in the regression model for non-transposed melodies. First, all the notes in the melodies were isochronous. Each note-onset overlapped exactly with each click of the rhythm. Hence the relative contrast between a strong beat (position 5) vs. a weak beat (position 4) may not have been enough for the listener to perceive, despite the use of physical accents. Second, instead of perceiving each note as an eighth note, listeners may have considered each note as a quarter note. In such a case, each melody spans two bars. Listeners may have considered the beat in position 4 to be a pick-up beat or anacrusis for the next successive bar (positions 5, 6, 7, 8). Therefore, position 4 may have been perceived as a position of strong rhythmic salience instead of weak rhythmic salience, thereby nullifying the role of rhythmic salience as a predictor. Third, all pitch changes of the altered note were still in the same scale as that of the standard melody. Therefore, although the change in tonality could be coded in the form of three levels of stability, none of the pitches were tonally unstable. This may have nullified the effect of tonal stability as a predictor.

In the case of transposed melodies, melodic contour and tonal stability were the only predictors that had explanatory power. Pitch distance and direction did not have any effect.

The effect of melodic contour on similarity judgments of transposed melodies has been highlighted in previous work by Dowling (1978). Listeners encode melodic structure at different levels. At a low level, encoding occurs at the level of pitches. This could be described as a form of local encoding. At a higher level, encoding may be occurring at the level of melodic contour where ups and downs in the melodic curve are captured. This could be described as a form of global encoding. When the comparison melody is transposed to a different key, listeners lose a basis for making direct pitch-to-pitch comparisons with the standard. However, higher-level features captured by melodic contour are still retained, thereby promoting their salience. Given that listeners cannot directly make pitch-related comparisons with the standard melody, as in the case of non-transposed melodies, tonal stability of pitches acts as a strong indicator of the scale of the melody.

IV. PERFORMANCE EVALUATION

Since Müllensiefen and Frieler (2004a) found that edit distance measurements with a rich symbolic representation compared well to human similarity judgments, we decided to evaluate our regression model, by comparing its performance with an edit distance method. Given that the comparison melodies in our stimuli differed from the standard melodies by one note, a string-matching edit distance method such as the one used by Wagner and Fischer (1974) would fail to detect any differences in similarity between the comparison melodies. All 12 comparison melodies would be considered as dissimilar by 1 unit from their standard melody. The Mongeau and Sankoff (1990) similarity-finding algorithm is an adaptation of the string-matching edit distance measure, used for comparing musical sequences across scores. While taking into account distances through insertions, deletions, and replacements of notes between two sequences, it also incorporates the consonance or dissonance of the replaced note within the dissimilarity measure. For example, if a note in the source sequence is replaced by another note, a weight is incorporated depending on the consonance or dissonance of the pitch interval between the original note and the new note, such that, the higher the dissonance, the greater the weight. Hence, we decided to compare the performance of our empirically derived measure (EDM) with the Mongeau and Sankoff measure (MSM).

A. Evaluation Experiment Method and Materials

For the purpose of testing our EDM, we created two standard test melodies, one in C major, and one in C minor. These melodies were of equal duration and consisted of eight isochronous notes, similar to the standard melodies used for deriving the regression model. We verified that these melodies were characteristic of real tonal melodies by correlation with the Krumhansl and Kessler (1982) profiles and resolution to the tonic. We created 12 comparison melodies for each standard by applying the exact procedure used for our study, and manipulating note 4 and note 5 of the standard.

Eight participants (3 female) from the Ryerson community participated in this experiment. They had 0 to 20 years of music training (M = 7.8, SD = 6.5) and ranged in age from 20 to 42 years (M = 28.8, SD = 8.2). Participants were administered the same experimental procedure as was used for the regression analysis. After listening to the comparison melody, they rated the level of similarity between the standard and comparison melodies on a Likert-type scale from 1 (least similar) to 5 (most similar).

We computed similarity values between the comparison melodies and their standards as predicted by the regression model in Equation (2). The MS algorithm was used to compute the dissimilarity values between the comparison melodies and their standards. Since the dissimilarity values range between 0 and 1, we converted these values into similarity values by subtracting them from 1. To compare the EDM and MSM with the mean similarity ratings of participants for each type of standard-variant comparison, we normalized similarity values in the following manner.

$$\textit{Normalized Similarity Value} = (SV-MinSV)/(MaxSV-MinSV) \qquad (4)$$

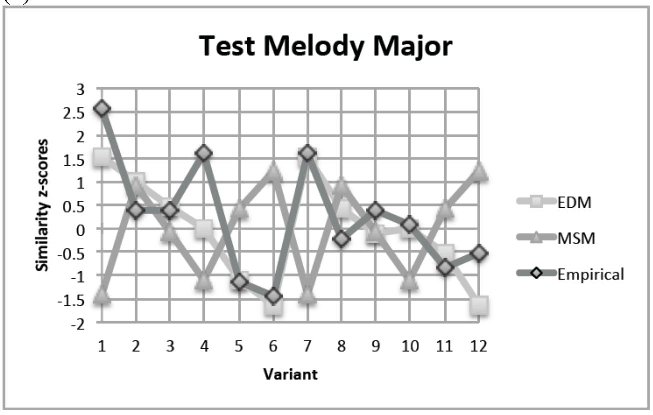
Here, SV is the similarity value, $MinSV$ is the minimum similarity value across all the comparison melodies, and $MaxSV$ is the maximum similarity value across all the comparison melodies. We calculated the mean total error for EDM by (a) computing the absolute difference between the normalized similarity values of the EDM and the normalized participant ratings, and (b) taking the average across all the values. Likewise, we calculated the mean total error for MSM.

Although averaging across all participants for each standard-variant comparison seems appropriate as a basis for developing our EDM, there are a few disadvantages. Each participant might have a substantially different judgment of similarity for a specific standard-variant comparison based on his or her subjective perception. An additional possibility is that a certain subset of participants might provide a high similarity rating while another subset of participants might provide a low similarity rating for the same standard-variant pair. The mean rating would fall somewhere in the middle of the scale, thus failing to capture the variance in participant ratings. To acknowledge these potential pitfalls of averaging, we also assessed the performance of EDM and MSM on the responses of all eight participants.

B. Evaluation Results and Discussion

The mean total error was 21.6% for the EDM and 52.5% for the MSM, indicating performance values of 78.4% and 47.5% respectively. The performance of the EDM for major and minor melodies was 82.8% and 74% respectively, whereas for the MSM it was 44.7% and 50.4% respectively. When compared with individual participant ratings, the performance of the EDM ranged from 59.5% to 76.5%, whereas the performance of the MSM ranged from 42.7% to 69.8%. These results clearly suggest that the EDM performed better than the MSM. One issue to be mindful of is that the MSM is used to detect similarities in musical sequences that differ in pitch as well as duration. The comparison melodies in our stimuli differ from their standards only in pitch but not in duration. Thus, the power of the MSM may have been compromised when applied to our stimuli. Figure 4 depicts participant similarity ratings alongside, EDM values, and MSM values, in the form of z-scores for all the 12 variant melodies, separated by major and minor scale melodies.

(a)



(b)

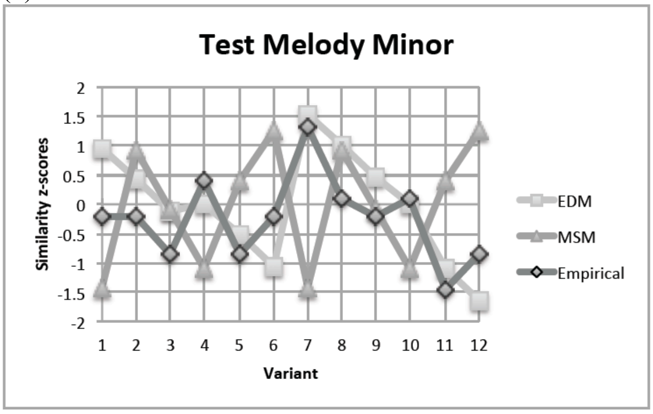


Figure 4. (a) Similarity comparison between empirical ratings, EDM values, and MSM values in major test melodies. (b) Similarity comparison between empirical ratings, EDM values, and MSM values in minor test melodies.

V. CONCLUSIONS AND FUTURE DIRECTIONS

While acknowledging previous work related to musical similarity that compared empirical judgments with algorithms, our intention for this study was to extend the available methodology within the context of music information retrieval. Specifically, our focus was on formulating an empirically derived measure for determining similarity between two melodies differing by one note. Through this study we developed such a measure on the basis of multiple linear regression analyses of similarity ratings. The predictors we evaluated were chosen because of their importance in the perception of melodic structure. Our findings indicate that for non-transposed melodies pitch distance, pitch direction, and to a certain degree melodic contour were important predictors that affected similarity judgments. For transposed melodies melodic contour and tonal stability significantly affected similarity judgments. We believe that rhythmic salience may not have had a significant effect because of possible ambiguities in the rhythmic structure (e.g., note 4 may have been perceived as a pick-up note).

The results of the regression models represent the first version of an empirically derived measure for determining the similarity between two melodies. In our next study, we intend to alter pitches on positions 3 (strong) and 6 (weak) instead of position 4 and 5. We also intend to incorporate melodies that differ by more than one note. This would require deriving a function that takes into account additional cognitive parameters such as the number of altered pitches, the primacy and recency effects of altered pitches, and the combined effect of musical predictors for each altered note.

ACKNOWLEDGMENT

This research was supported by a Mitacs Elevate postdoctoral fellowship awarded to Naresh N. Vempala, co-sponsored by Mitacs and *wave*DNA, Inc. We are grateful to Glen Kappel at *wave*DNA for providing us with valuable comments and suggestions. We would like to thank Geraint Wiggins for providing us with useful insights on similarity-finding methods.

REFERENCES

Clausen, M., & Kurth, F. (2002). A Unified Approach to Content-based and Fault-tolerant Music Identification. In *Second International Conference On Web Delivering of Music*, pp. 0056. Darmstadt, Germany.

Dowling, W. J. (1972). Recognition of melodic transformations: Inversion, Retrograde, and retrograde Inversion. *Perception and Psychophysics*, *12*, 417-421.

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

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.

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

Frances, R. (1958). *La perception de la musique*. Paris: Vrin.

Kim, K., Lee, D., Yoon, T., & Lee, J. (2008). A Music Recommendation System based on Personal Preference Analysis. In *Proceedings of the First International Conference on the Applications of Digital Information and Web Technologies*, 102-106.

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

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.

McNab, R. J., Smith, L. A., Witten, I. H., Henderson, C. L., & Cunningham, S. J. (1996). Towards the digital music library: Tune retrieval from acoustic input. In *Proceedings of the 1st ACM International Conference on Digital Libraries*, 11-18. Bethesda, MD: ACM Press.

Mongeau, M., & Sankoff, D. (1990). Comparison of musical sequences. *Computers and the Humanities, 24*, 161-175.

Müllensiefen, D., & Frieler, K. (2004a). Measuring Melodic Similarity: Human vs. Algorithmic Judgments. In *Proceedings of the Conference on Interdisciplinary Musicology*. Austria: Graz.

Müllensiefen, D., & Frieler, K. (2004b). Cognitive adequacy in the measurement of melodic similarity: Algorithmic vs. human judgments. *Computing in Musicology, 13*, 147-176.

Müllensiefen, D., & Frieler, K. (2007). Modelling experts' notions of melodic similarity. *Musicae Scientiae*, Discussion Forum 4A, 183-210.

Murdock Jr., B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*, 482-488.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The Implication-Realization model*, University of Chicago Press, Chicago.

Pauws, S. (2002). CubyHum: A Fully Operational Query-by-humming System. In *Proceedings of the 3rd International Conference on Music Information Retrieval*, 187–196.

Prechelt, L., & Typke, R. (2001). An interface for melody input. *ACM Transactions on Computer-Human Interaction*, *8*, 133-149, 2001.

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

Schubert, E., & Stevens, C. (2006). The effect of implied harmony, contour and musical expertise on judgments of similarity of familiar melodies. *Journal of New Music Research*, *35*, 161-174.

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.

Typke, R., Wiering, F., & Veltkamp, R. C. (2005). A Survey of Music Information Retrieval Systems. In *Proceedings of the 6th International Conference on Music Information Retrieval*, 153-160.

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

Wagner, R., & Fischer, M. (1974). The string-to-string correction problem. *Journal of the ACM*, *21*, 168-173, 1974.