

Measuring Musical Rhythm Similarity: Further Experiments with the Many-to-Many Minimum-Weight Matching Distance

Godfried T. Toussaint¹, Seung Man Oh²

¹Computer Science Program, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

²Freelance IT Developer, Seoul, Korea

Email: gt42@nyu.edu

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

Received: September 24, 2016

Accepted: November 25, 2016

Published: November 28, 2016

Abstract

Musical rhythms are represented as sequences of symbols. The sequences may be composed of binary symbols denoting either silent or monophonic sounded pulses, or ternary symbols denoting silent pulses and two types of sounded pulses made up of low-pitched (*dum*) and high-pitched (*tak*) sounds. Experiments are described that compare the effectiveness of the many-to-many minimum-weight matching between two sequences to serve as a measure of similarity that correlates well with human judgements of rhythm similarity. This measure is also compared to the often used *edit* distance and to the one-to-one minimum-weight matching. New results are reported from experiments performed with three widely different datasets of real-world and artificially generated musical rhythms (including Afro-Cuban rhythms), and compared with results previously reported with a dataset of Middle Eastern *dum-tak* rhythms.

Keywords

Musical Rhythms, Rhythm Similarity Measures, Many-to-Many Minimum-Weight Matching, Edit Distance, One-to-One Minimum-Weight Matching, Rhythm Perception, Afro-Cuban Rhythms, Middle Eastern Rhythms

1. Introduction

Measuring the similarity between musical rhythms is a problem of paramount relevance in the areas of music information retrieval [1], musicology [2], phylogenetic analysis [3] [10], and music perception [4]. The approaches taken to modelling measures of rhythm similarity vary according to whether music is represented acoustically [5] or

symbolically [6] [9]. In the present study, musical rhythms are represented symbolically as 2-symbol sequences, in which each symbol denotes monophonic sounds (onsets) or silences (rests) of unit time duration. In this report experiments are described that compare the effectiveness of the many-to-many minimum-weight matching between two sequences to serve as a measure of similarity that correlates well with human judgements of rhythm similarity across widely different real-world and artificially generated datasets of rhythms, including Afro-Cuban and Middle Eastern rhythms. The many-to-many minimum-weight matching distance is also compared to the often used *edit* distance and the one-to-one minimum-weight matching distance. The present study extends previous work carried out with only one dataset consisting of Middle Eastern *dum-tak* rhythms [8], with the goal of testing the generalizability of those results to other widely different datasets.

2. The Distance Measures Compared

2.1. The Many-to-Many Minimum-Weight Matching Distance

A powerful measure of the distance (dissimilarity) between two sequences of symbols (elements) is the many-to-many minimum-weight matching distance [7]. The calculation of this distance measure is illustrated using box notation in **Figure 1**, where sequence *A* has four sounded pulses (filled boxes) and sequence *B* has 5 sounded pulses. Empty boxes denote rests, and each box represents one unit of time. What distinguishes the many-to-many matching from other matching schemes such as the one-to-one matching or the many-to-one matching is that the many-to-many matching allows several sounded pulses to be assigned to a single sounded pulse from either sequence to the other. In the example of **Figure 1** the sounded pulses numbered 1 and 3 in sequence *A* are assigned to sounded pulse number 2 of sequence *B*, and the three sounded pulses in sequence *B* (in positions 12, 14 and 16) are assigned to sounded pulse 14 in sequence *A*. The cost of each assignment is equal to the number of pulses that separate the source pulse from the target pulse, and the many-to-many distance is the sum of the individual costs. Thus in **Figure 1** this distance equals $2 + 0 + 1 + 0 + 2 = 5$. The many-to-many minimum-weight matching distance was computed with the Hungarian algorithm [15] as implemented by Munkres [16].

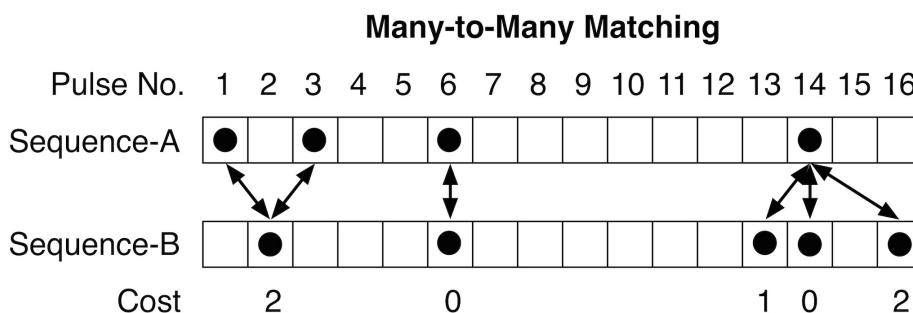


Figure 1. Illustration of the many-to-many minimum-weight matching distance.

2.2. The Edit Distance

The edit distance between two sequences is defined as the minimum number of insertions, deletions, and substitutions of symbols required to transform one sequence to the other [2] [4]. An insertion operation inserts one symbol (either silent or sounded pulse) resulting in a rhythm of longer duration. A deletion operation deletes one symbol, shortening the duration of the rhythm. A substitution operation substitutes one symbol for another, leaving the duration of the rhythm unaltered. Consider for example the calculation of the edit distance between the two 16-pulse sequences A and B in **Figure 2**. In the first edit operation A is converted to the 15-pulse sequence A^* by deleting silent pulse number 5. In the second operation the 15-pulse sequence A^* is converted into the 15-pulse sequence A^{**} by substituting the sounded pulse number 12 by a silent pulse. Finally, the 15-pulse sequence A^{**} is converted to the 16-pulse sequence B by inserting a sounded pulse between pulses 13 and 14, for a total of 3 operations. No other set of operations to convert A to B uses less operations, and therefore the edit distance between A and B is equal to 3. The edit distance was computed using dynamic programming [8].

2.3. The One-to-One Minimum-Weight Matching Distance

The one-to-one minimum-weight matching distance between two rhythms that contain an unequal number of sounded pulses assigns every sounded pulse of the denser rhythm to a unique sounded pulse of the sparser rhythm, thus leaving some sounded pulses in the denser rhythm unassigned, in such a way that the resulting overall weight (cost) is minimized over all such possible assignments [14]. For example, the two rhythms illustrated in **Figure 3** contain unequal numbers of pulses (8 versus 7) and sounded pulses (5 versus 3). Therefore, two sounded pulses in the denser upper rhythm (2 and 7) are left unassigned, and three of them are selected that yield a minimum cost

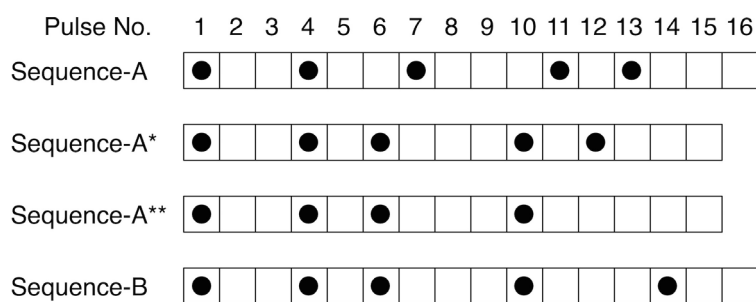


Figure 2. Illustration of the calculation of the edit distance.

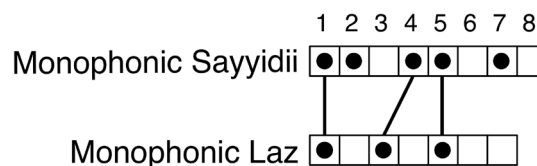


Figure 3. Illustration of the one-to-one minimum-weight matching distance.

matching distance equal to 1 (the distance between pulse 4 in the upper rhythm and pulse 3 in the lower rhythm). Like the many-to-many minimum-weight matching distance, the one-to-one version was computed with the Hungarian algorithm [15].

3. Datasets Used

Three datasets of rhythms were used in the experiments reported here. The first dataset comprises a group of Afro-Cuban rhythms studied by Rey [12] in an evolutionary context, for which a phylogenetic analysis subsequently carried out failed to support Rey's ethnographic account [13]. These rhythms are shown in box notation in **Figure 4**. The rhythms used by Rey comprise the first seven in the list. The second rhythm called *tresillo* in Cuba has Inter-Onset-Intervals 3-3-2, and is used all over the world. Its two rotations 2-3-2 and 3-2-3 (appended at the bottom of the list) were also included in this dataset, since they are also used in many cultures around the world.

The second dataset consists of 17 rhythms, one of which is the clave son [17], and the remaining 16 are mutations of the clave son obtained by either substituting one of its 5 sounded pulses with a silent pulse (rest) or replacing one of its 11 silent pulses with a sounded one [4]. This dataset is shown in **Figure 5**. The labels indicate the type of mutation made to the clave son. The notation D- i indicates that the sounded pulse at position i was replaced by a silent pulse (deletion of a sound). The notation I- i indicates that the silent pulse at position i was replaced by a sounded pulse (insertion of a sound). Although this is an artificially generated set, since the rhythms consist of minimal deletions and insertions of sounded pulses, some of the rhythms are actually heard in practice. In particular, it is not uncommon to hear the rhythm D-13, with the last sounded pulse of the clave son missing, resulting in a more syncopated rhythm, and the rhythms I-5 and I-6, in which the second and third sounded pulses of the clave son are embellished with an additional sounded pulses either just before or just after a sounded pulse, respectively.

The third dataset, shown in **Figure 6**, is made up of the clave son rhythm and 12 permutations of its inter-onset intervals (IOIs) that were selected at random from all possible permutations [4]. The IOI is standard terminology for the distance between two adjacent sounded pulses. This artificially generated dataset also happens to include one rhythm frequently heard in practice, namely, 43,234 common in rap music [8].

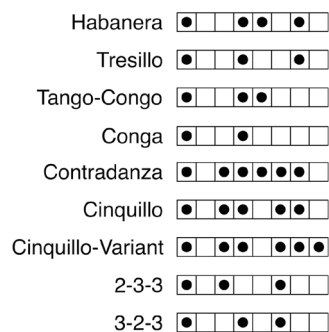


Figure 4. The Afro-Cuban rhythms studied by Rey [12] plus two rotations of the *tresillo*.

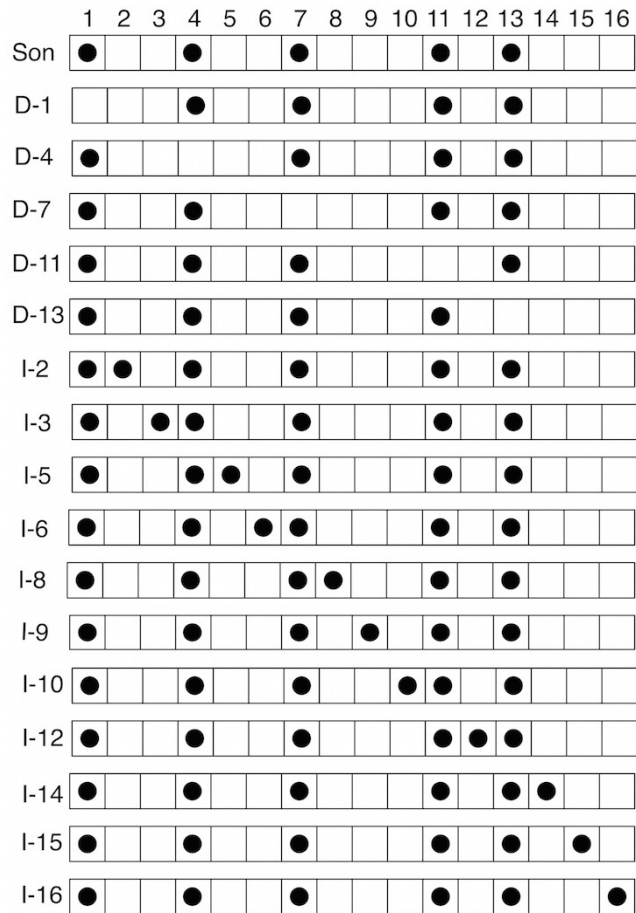


Figure 5. The clave son and 16 mutations obtained by substitutions of sounded and silent pulses.

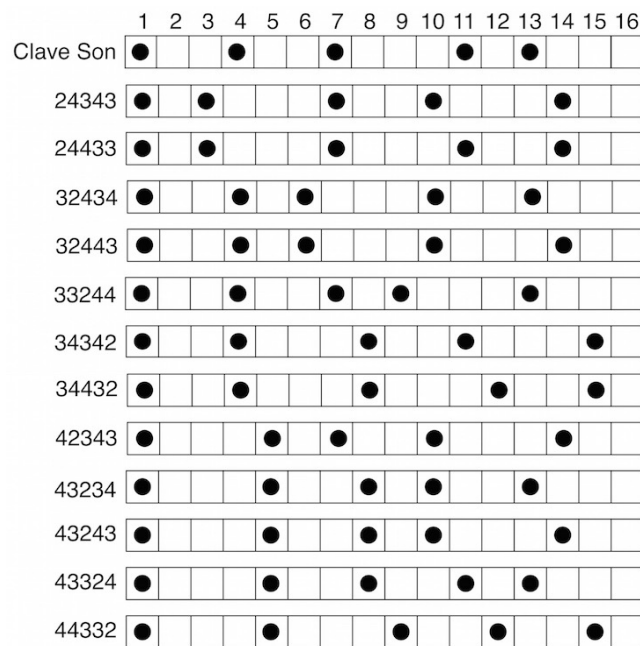


Figure 6. The clave son and 12 permutations of its IOIs 33424.

The results with the three preceding datasets were compared with the results obtained previously with the dataset consisting of the nine Middle Eastern *dum-tak* rhythms [11] shown in **Figure 7**.

4. Results

The results of the experiments are listed in **Table 1**, which lists the Spearman rank correlation coefficients (with significance tests in the form of *p*-values) for the many-to-many minimum-weight matching distance, compared with the edit distance and the one-to-one minimum-weight matching distance, for all four datasets. The listening tests with human subjects for all four datasets were carried out at Harvard University [4] [11] [13]. A description of the experiments will not be duplicated here, and the reader is referred to those references for more detailed information. The correlations for the *dum-tak* rhythms between the many-to-many matching distance and the human subjects and edit distance were reported in the recent paper by Toussaint and Oh [8].

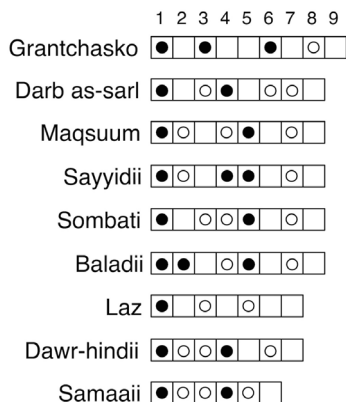


Figure 7. The Middle Eastern *dum-tak* rhythms [11].

Table 1. Spearman rank correlations for the many-to-many minimum-weight matching.

Correlation with Human Subjects			
Rey Rhythms	17 Son Mutations	13 Son IOI Permutations	Dum-Tak Rhythms [8]
0.322	0.349	0.389	0.661
<i>p</i> = 0.0196	<i>p</i> = 0.0011	<i>p</i> = 0.0013	<i>p</i> = 0.0006
Correlation with Edit Distance			
Rey Rhythms	17 Son Mutations	13 Son IOI Permutations	Dum-Tak Rhythms [8]
0.766	0.378	0.384	0.566
<i>p</i> = 0.0001	<i>p</i> = 0.0598	<i>p</i> = 0.0009	<i>p</i> = 0.0076
Correlation with One-to-One Matching Distance			
Rey Rhythms	17 Son Mutations	13 Son IOI Permutations	Dum-Tak Rhythms
-0.142	0.839	1.000	0.758
<i>p</i> = 0.2352	<i>p</i> = 0.0001	<i>p</i> = 0.0001	<i>p</i> = 0.0001

5. Discussion and Conclusion

The main goals of the experiments carried out in this research project were to determine how the many-to-many minimum-weight matching distance predicts human judgments of rhythm similarity across groups of rhythms that differ greatly from each other in terms of genre, cycle length (duration), and the distribution of sounded pulses within the rhythmic cycles, as well as to ascertain how similar the many-to-many matching distance measure is to two other popular measures of rhythm dissimilarity, the edit distance and the one-to-one minimum-weight matching distance.

Regarding how well the many-to-many distance predicts human judgements, although the correlations with the three new data sets are lower than that obtained for the *dum-tak* rhythms (0.322, 0.349, and 0.389 for the first three datasets versus 0.661 for the *dum-tak* rhythms, as shown in **Table 1**) all are moderately correlated and highly statistically significant. However, previously published results show that the edit distance significantly outperforms the many-to-many distance for these three datasets as evidenced by the results in **Table 2**. The correlations between the many-to-many distance and the human judgments in **Table 1**, averaged over the four datasets, is 0.430 with standard deviation 0.156, whereas for the edit distance in **Table 2** the average correlation is 0.741 with standard deviation 0.127. Thus although the many-to-many distance is robust across these four datasets, so is the edit distance. Furthermore, the edit distance appears to be a more consistent and accurate predictor of human judgements of perceived rhythm similarity than the many-to-many matching distance.

Concerning the similarity between the many-to-many matching distance and the edit distance, the four datasets may be categorized into two distinct groups: real-world rhythms (Afro-Cuban and Middle Eastern) and artificially generated rhythms (mutations of sounded pulses and permutations of IOIs of the clave *son*). These two groups of datasets can be distinguished mainly by two features: their total number of pulses, and the variability in their number of sounded pulses. The real-world rhythms have lengths that vary from 6 to 9 pulses, whereas the artificial rhythms all have 16 pulses. Furthermore, the number of sounded pulses in the real-world rhythms varies between 2 and 6 (out of a maximum of 9 pulses), whereas the artificial rhythms all have between 4 and 6 sounded pulses (out of a maximum of 16 pulses). The results in **Table 1** show that the two measures correlate more highly with each other for the real-world datasets than for the artificial datasets. The correlations for the Rey Rhythms and the *dum-tak* rhythms are 0.766 and 0.566, respectively, whereas for the 17 *son* mutations and 13 *son* IOI permutations the correlations are 0.378 and 0.384, respectively. The fraction of the pulses that are sounded for the real-world rhythms is 62.6%, whereas the corresponding fraction for the artificially generated rhythms is 32.5%. Therefore, in these datasets the

Table 2. Correlations between the edit distance and human judgments.

Rey Rhythms [13]	17 Son Mutations [4]	13 Son IOI Permutations [4]	Dum-Tak Rhythms [8]
0.594	0.866	0.828	0.677
$p = 0.0002$	$p = 0.0001$	$p = 0.0001$	$p = 0.001$

real-world rhythms are about twice as dense as the artificially generated rhythms. This suggests the hypothesis that the many-to-many matching and edit distances tend to behave in a similar manner for dense rhythms, and tend to differ from each other for sparse rhythms. It would be interesting to test this hypothesis further with other data sets.

With respect to the one-to-one minimum-weight matching distance [14], the correlations between it and the many-to-many matching distance differ widely across the four datasets, as evidenced in **Table 1**. For two rhythms that have the same number of sounded pulses the many-to-many matching distance is equivalent to the one-to-one matching distance. Therefore, for such rhythms the correlation between the two measures should be 1.0. This is confirmed with the dataset consisting of the IOI permutations of the clave son since they all have 5 onsets. Thus, for rhythms that contain an almost equal number of sounded pulses the correlation between the measures is expected to be high. This is also confirmed with the dataset consisting of mutations which delete or insert only one sounded pulse, as well as the Middle Eastern *dum-tak* rhythms, almost all of which have five sounded pulses. The dataset for which the two measures do not correlate at all consists of the Afro-Cuban rhythms in Rey's dataset. This may be explained by the fact that these rhythms exhibit a high variability in their number of sounded pulses ranging from 2 to 6. This variability creates significant differences in the scores obtained with each measure. For example, the conga and contradanza rhythms have 2 and 6 sounded pulses, respectively. These two rhythms are considerably different from each other, and the many-to-many matching distance is 7, reflecting this difference. On the other hand, the one-to-one matching distance between them is 0, and completely ignores the difference.

Returning to the goal of predicting accurately the human judgements of rhythm similarity, the many-to-many matching distance is not able to compete with the edit distance for these four datasets (see **Table 1** and **Table 2**). However, discarding the former distance outright based on the experiments reported here is premature. It may be possible, at least for the case of rhythms that contain polyphonic sounded pulses such as *dum-tak* rhythms, to obtain better results with a modification of the many-to-many matching distance that restricts the assignments between the sounded pulses of two rhythms in such a way that only sounds of the same type are assigned to each other. Thus the resulting bipartite graph between two *dum-tak* rhythms would not contain edges between a *dum* of the first rhythm and a *tak* of the second rhythm. Experiments with such a modification of the many-to-many minimum-weight matching distance are planned for future investigation.

Acknowledgements

This research was supported by a grant from the Provost Fund at New York University Abu Dhabi and a Research Enhancement Fund grant from the NYUAD Institute, for the project titled: *Cross-Disciplinary and Multi-Cultural Perspectives on Musical Rhythm*.

References

- [1] Panteli, M., Bogaards, N. and Honingh, A. (2014) Modeling Rhythm Similarity for Electronic Dance Music. *Proceedings of the 15th International Society for Music Information Retrieval Conference*, Taipei, Taiwan, 27-31 October, 537-542.
- [2] Orpen, K.S. and Huron, D. (1992) Measurement of Similarity in Music: A Quantitative Approach for Non-Parametric Representations. *Computers in Music Research*, **4**, 1-44.
- [3] Toussaint, G.T. (2016) Phylogenetic Analysis of the Ancient Greek Paeonic Rhythms. *Proceedings of Bridges Finland 2016: Mathematics, Music, Art, Architecture, Education, Culture*, Jyväskylä, 9-13 August 2016, 363-366.
- [4] Post, O. and Toussaint, G.T. (2011) The Edit Distance as a Measure of Perceived Rhythmic Similarity. *Empirical Musicology Review*, **6**, 164-179.
- [5] Berenzweig, A., Logan, B., Ellis, D.P.W. and Whitman, B. (2004) A Large-Scale Evaluation of Acoustic and Subjective Music-Similarity Measures. *Computer Music Journal*, **28**, 63-76. <http://dx.doi.org/10.1162/014892604323112257>
- [6] Beltran, J.F., Liu, X., Mohanchandra, N. and Toussaint, G.T. (2015) Measuring Musical Rhythm Similarity: Statistical Features versus Transformation Methods. *Journal of Pattern Recognition and Artificial Intelligence*, **29**, 1-23. <http://dx.doi.org/10.1142/s0218001415500093>
- [7] Mohamad, M., Rappaport, D. and Toussaint, G.T. (2015) Minimum Many-to-Many Matchings for Computing the Distance Between Two Sequences. *Graphs and Combinatorics*, **31**, 1637-1648. <http://dx.doi.org/10.1007/s00373-014-1467-4>
- [8] Toussaint, G.T. and Oh, S.M. (2016) Measuring Musical Rhythm Similarity: Edit Distance versus Minimum-Weight Many-to-Many Matchings. *Proceedings of the 16th International Conference on Artificial Intelligence*, Las Vegas, 25-28 July 2016, 186-189.
- [9] Toussaint, G.T. (2013) *The Geometry of Musical Rhythm: What Makes a "Good" Rhythm Good?* Chapman & Hall/CRC Press, Boca Raton.
- [10] Toussaint, G.T. (2015) Modeling Musical Rhythm Mutations with Geometric Quantization. In: Melnik, R., Ed., *Mathematical and Computational Modeling: With Applications in Natural and Social Sciences, Engineering, and the Arts*, John Wiley & Sons, Inc., 299-308. <http://dx.doi.org/10.1002/9781118853887.ch12>
- [11] Toussaint, G. T., Campbell, M. and Brown, N. (2011) Computational Models of Symbolic Rhythm Similarity: Correlation with Human Judgments. *Analytical Approaches to World Music Journal*, **1**, 380-430.
- [12] Rey, M. (2006) The Rhythmic Component of *Afrocubanismo* in the Art Music of Cuba. *Black Music Research Journal*, **26**, 181-212.
- [13] Toussaint, G. T., Matthews, L., Campbell, M. and Brown, N. (2012) Measuring Musical Rhythm Similarity: Transformation versus Feature-Based Methods. *Journal of Interdisciplinary Music Studies*, **6**, 23-53.
- [14] Karp, R.M. and Li, S.Y.-R. (1975) Two Special Cases of the Assignment Problem. *Discrete Mathematics*, **13**, 129-142. [http://dx.doi.org/10.1016/0012-365X\(75\)90014-X](http://dx.doi.org/10.1016/0012-365X(75)90014-X)
- [15] Kuhn, H.W. (1955) The Hungarian Method for the Assignment Problem. *Naval Logistics Quarterly*, **2**, 83-97. <http://dx.doi.org/10.1002/nav.3800020109>
- [16] Munkres, J. (1957) Algorithms for the Assignment and Transportation Problems. *Journal of the Society of Industrial and Applied Mathematics*, **5**, 109-133. <http://dx.doi.org/10.1137/0105003>
- [17] Toussaint, G.T. (2011) The Rhythm that Conquered the World: What Makes a "Good" Rhythm Good? *Percussive Notes*, November, 52-59.