Clustering and classification of music using interval categories

Honingh, A.K.; Bod, L.W.M.

Published in:
Mathematics and Computation in Music

DOI:
10.1007/978-3-642-21590-2_30

Clustering and classification of music by interval categories

Aline Honingh and Rens Bod

Institute for Logic, Language and Computation, University of Amsterdam
A.K.Honingh@uva.nl, Rens.Bod@uva.nl

Abstract. We present a novel approach to clustering and classification of music, based on the concept of interval categories. Six interval categories exist, each with its own musical character. A piece of music can be represented by six numbers, reflecting the percentages of occurrences of each interval category. A piece of music can, in this way, be visualized as a point in a six dimensional space. The three most significant dimensions are chosen from these six. Using this approach, a successful visual clustering of music is possible for 1) composers through various musical time periods, and 2) the three periods of Beethoven, which illustrates the use of our approach on both a general and a specific level. Furthermore, we will see that automatic classification between tonal and atonal music can be achieved.

Keywords: interval category, pitch class set, classification, clustering

1 Introduction

Comparing music can be accomplished on several levels. Finding similarities and dissimilarities is important for the development of many music informationretrieval systems. Since the amount of digitized music has increased enormously over the past few years, the need for tools to organize, order and cluster music has increased as well. Automatic music classification has been investigated on the basis of different ideas, such as compression distance [2], Hidden Markov Models [1], and n-gram models [3].

In this paper we present a new approach to music clustering and classification on the basis of interval categories (IC's). With this new approach, we can not only cluster and classify music according to certain genres or styles, but we may also be able to explain the differences and commonalities in terms of interval and pitch content.

2 Interval categories

It has been shown that pitch class sets (pc-sets) can be grouped into six interval categories [7, 5]. Every interval category is associated with an interval class. IC1 corresponds to intervals of 1 (or 11), IC2 corresponds to intervals of 2 (or 10),

IC3 corresponds to intervals of 3 (or 9), and so on, meaning that a pitch class set belonging to ICn contains a dominating number of intervals n. Every pitch class set can be classified into one of these categories, and it will belong to the category it is most similar to. For example, pitch class sets $\{0, 1, 2\}$, $\{0, 2, 3, 4, 5\}$, $\{0, 1, 2, 3, 5\}$ will all be grouped into IC1 since the number of semitones in each set dominates the set.

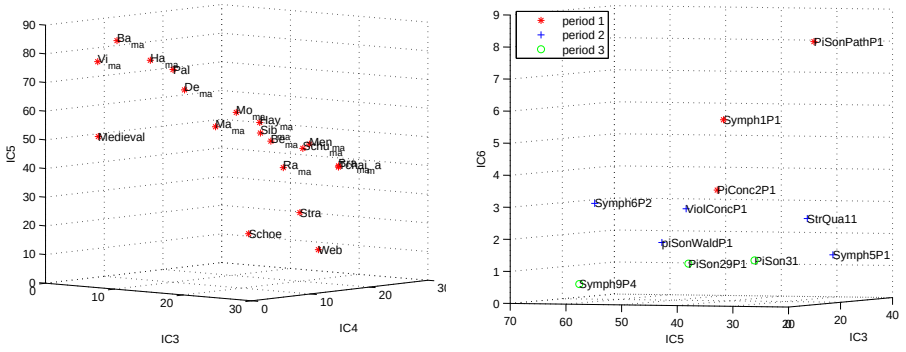
An entire piece or corpus of music can be represented by a distribution of interval categories. A piece of music is segmented and the notes in every segment form a pc-set. Of every pc-set it can be calculated to which IC it belongs. In this way, a piece of music can be represented by its IC distribution, listing for each category the percentage of occurrence in the piece or corpus. For a more elaborate explanation of these interval categories, as well as the detailed method to represent a piece of music in terms of interval categories, see [5, 4].

3 Composer clustering

Western art music, also referred to as classical music, is generally grouped into several musical periods (listed in table 1) and is known to express different characteristics for every period. Using the IC distributions, we have visualized several composers and hope to see that the various musical periods can be distinguished in clusters. The composers we have used are listed in table 1. For each composer we have selected around five pieces of music. The Medieval period was treated as one item, since for most medieval music the composers are unknown. Since the musical mode has been shown to influence the IC distribution [4], we have chosen to only use music in major mode for the present visualization purpose. Each composer is represented by an IC distribution that was calculated using the selected music. The six interval categories give rise to six dimensions, but we only chose three of these (IC3, IC4 and IC5) to visualize the composers, see figure 1a. The Baroque music as well as the atonal music (represented by composers Schoenberg and Webern) are found as separate clusters, while the Romantic and Classical music are difficult to distinguish in the figure.

Table 1. Periods of European art music and the composers we used for figure 1a.

Period	**date**	**composers (abbreviation)**
Medieval	500-1400	
Renaissance	1400-1600	Palestrina (Pal)
Baroque	1600-1750	Bach (Ba), Händel (Ha), Vivaldi (Vi)
Classical	1750-1830	Haydn (Hay), Mozart (Mo), Beethoven (Be), Schubert (Schu)
Romantic	1830-1900	Brahms (Bra), Mahler (Ma), Tchaikovsky (Tchai), Debussy (De), Mendelssohn (Men), Sibelius (Sib)
Modern	1900–...	Ravel (Ra), Stravinsky (Stra), Schoenberg (Schoe), Webern (Web)



(a) (b)

Fig. 1. a) 3-D plot of music from composers listed in table 1. The label 'ma' indicates the major mode. b) 3-D plot of music from the three periods of Beethoven.

4 Tonal - atonal classification

On the basis of the results above, we have tried to automatically separate tonal from atonal music. For every piece of music, if the ratio category 5/category 1 was higher than a certain threshold, the piece was classified as tonal, and otherwise as atonal. To evaluate our results, we compared them to the results of a baseline algorithm, which reads: for every bar, the diatonic scale is selected that matches most of the notes in the bar, after which the number of notes that are not elements of that scale are counted. An average of those 'atonal notes' is calculated over the whole piece. If this number is higher than a certain threshold, the piece is classified as atonal and otherwise as tonal.

The results can be found in table 2. We conclude that the overall classification on the basis of IC's is quite successful (94.7 % correct).

Table 2. Results of classification of tonal and atonal music.

	Correctly classified **atonal** pieces	Correctly classified **tonal** pieces	Total number of correctly classified pieces
IC algorithm	19	53	72 (94.7 %)
baseline algorithm	14	53	67 (88.2 %)
total number of pieces	20	56	76

5 Beethoven's three periods

Beethovens compositional career is usually divided into three periods [6]. From each period, we have chosen a number of works, and calculated the IC distribu-

tion of each piece. From period 1 we chose the piano sonata Pathetique part 1 (PiSonPathP1), piano concerto 2 part 1 (PiConc2P1), and symphony 1 part 1 (Symph1P1). From period 2, we chose symphony 5 part 1 (Symph5P1), piano sonata Waldstein part 1 (PiSonWaldP1), string quartet no 11 (StrQuar11), violin concerto part 1 (ViolConcP1), and symphony 6 part 2 (Symph6P2). From period 3, we chose symphony 9 part 4 (Symph9P4), piano sonata 29 part 1 (PiSon29P1), and piano sonata 31 (PiSon31).

IC3, IC5 and IC6 have been chosen as dimensions to visualize the pieces in, see figure 1b. We see that the pieces from the three different periods can be distinguished on the basis of dimension IC6. Figure 1b surprisingly shows that Beethoven has been using relatively many tritones (IC6) in his early works, and relatively few tritones in his late works. Of course, these differences in tritones do not fully characterize the three periods. However, the fact that a visual classification on the basis of this category is possible, suggests that the use of tritones contributes to the well-known classification of Beethoven's music.

6 Conclusions

We have presented a new approach to algorithmically cluster music based on interval categories (IC's). We have seen that, with this approach, 1) composers can be clustered according to the musical period they belong to, 2) tonal music can be automatically distinguished from atonal music, and 3) music from Beethoven can be clustered according to Beethoven's three different periods.

We should keep in mind that these IC clusterings are only based on pitch intervals and therefore can never capture all of the differences and commonalities of the investigated music. However, it is striking that a classification *can* be made with the use of this simple method and thus the approach provides a simple and powerful representation of music that can be used to investigate musical similarity on the basis of tonalness, composer, musical period and possibly more.

References

1. Chai, W., Vercoe, B.: Folk music classification using hidden Markov models. In Proceedings of ICAI (2001)
2. Cilibrasi, R., Vitanyi, P., De Wolf, R.: Algorithmic Clustering of Music Based on String Compression. Comput. Music J. 28(4),49-67 (2004)
3. Hillewaere, R., Manderick, R., Conklin, D.: String quartet classification with monophonic models. In Proceedings of ISMIR (2010). Utrecht, the Netherlands
4. Honingh, A.K., Bod, R.: Pitch class set categories as analysis tools for degrees of tonality. In Proceedings of ISMIR (2010). Utrecht, the Netherlands
5. Honingh, A.K., Weyde, T., Conklin, D.: Sequential association rules in atonal music. In Proceedings of MCM (2009). New Haven, USA
6. Kerman, J. et al.: Beethoven, Ludwig van. In Grove Music Online. Oxford Music Online. `http://www.oxfordmusiconline.com/subscriber/article/grove/music/40026pg11` accessed January 11, 2011
7. Quinn, I.: Listening to similarity relations. Perspect. New Music 39, 108-158 (2001)