# Symbolic Music Genre Classification based on Note Pitch and Duration*

Ioannis Karydis

Aristotle University of Thessaloniki, Thessaloniki 54124, Greece
karydis@delab.csd.auth.gr

**Abstract.** This paper presents a music genre classification system that relies on note pitch and duration features, derived from their respective histograms. Feature histograms provide a simple but yet effective classifier for the purposes of genre classification in intra-classical genres such as sonatas, fugues, mazurkas, etc. Detailed experimental results illustrate the significant performance gains due to the proposed features, compared to existing baseline features.
**Keywords:** Music genre classification, music features, histograms, pitch, duration, content-based information retrieval.

## 1 Introduction

Digitised music exists in broadly two categories depending on whether its recording contains directions of what to be played by a performer or a particular audio-recorded performance of a piece. The former representation of music is the symbolic, while the latter is the acoustic. Music Information Retrieval (MIR) is accordingly divided into two categories depending on the representation of music that is under examination.

Although young a field, MIR and especially Content-Based MIR (CBMIR) mainly orientate towards acoustic data, a fact that can easily be partially explained by the popularity of acoustic recordings. Though, the two representations are interconnected with acoustic music being, improvisation set aside, up to a great degree the product of symbolic music. Thus, taking into consideration the relationship between the two representations of music and the existence of very large acoustic databases (for both commercial and not purposes) one can imagine not only the existence of large analogous collections of symbolic music but also the significance of MIR on symbolic data, especially for music distribution.

One of the necessities that prevails in MIR is genre classification. Apart from the obvious significance to numerous occupations (retailers, librarians, musicologists, e.t.c.) as a means for music organisation, genre classification is additionally important as research indicates that liking a music piece can adhere to the performance style instead of the actual piece itself [4]. Since predefined metadata in symbolic music data are rare and their manual appointment inhibits difficulties

* This research is supported by the *ΗΡΑΚΛΕΙΤΟΣ* and *ΠΥΘΑΓΟΡΑΣ ΙΙ* national programs funded by *ΕΠΕΑΕΚ*.

and potential inconsistencies, the need for an effective automatic means of music classification unfolds as the collections of symbolic digital music files increase at a rapid rate. Moreover, genre classification is of great assistance to the wide public accessing musical archives, offering increased ease in identifying types of music.

As aforementioned, musical pieces in symbolic format represent the intention of the composer towards the performer. Thus, the symbolic representation engulfs an excess of information, that may not always be perceivable in the respective acoustic piece. In order to process all the information included in the music files, one can rely on perceptual criteria (features) related to pitch, rhythm, timbre, etc of the music in order to characterise a musical genre. The key to success is the choice of features to be based upon, while the effectiveness of the classifier, although still important, remains secondary as is limited by the feature selectivity. In this paper, we focus on the note pitch and duration information of the musical data.

### 1.1 Contribution and Paper Organisation

This paper examines the problem of determining the musical genre of a musical piece, provided in symbolic representation. Based on prior work on symbolic music genre classification, we focus on note pitch and duration information of the musical data.

Current research on music genre classification based on musical feature histograms [5], has been isolated on the pitch information of notes solely. Although pitch is described in the literature as one of the predominant musical characteristics [2], rhythm, one of the main dimensions of which is note durations, is also given high credit and current research is not considering it.

Moreover, current research examines broad categories that, although up to a degree overlapping and vague, present far more distinctive characteristics than the sub-categories of any category.

To address these issues, this paper proposes the following:

- Re-examination of the selectivity of pitch features in different music categories that present more similarities,
- Introduction of duration and pitch-duration combination features that are based on pitch and duration information of the notes,
- A differentiated approach to pitch feature as described by current research.

The rest of the paper is organised as follows. Section 2 is devoted in background information and related research as far as symbolic music genre classification is concerned. Moreover, a baseline approach is reviewed therein and the motivating factors that led to this research are summarised. Extending the idea proposed in Section 2.1, Section 3 provides a complete account of the features proposed in this paper. Subsequently, Section 4 presents and discusses the experimentation results obtained. Finally, the paper is concluded by a summary and the intended future work in Section 5.

## 2 Background and Related Work

Musical genre classification is one of the key areas MIR researchers are interested in. Although, as already mentioned, genre classification research is mainly oriented towards acoustic data, approaches for symbolic data do exist and have interesting results to demonstrate.

Tzanetakis et al. [5] presented pitch histograms as a way to represent the pitch content of music signals both in symbolic and acoustic form. Based on features extracted from these histograms the authors of [5] managed a 50% accuracy for 5 genres (for more details see Section 2.1).

Following the participation success of the International Symposium on Music Information Retrieval (ISMIR) conference on 2005, the MIREX competition was held. The goal of the contest was to classify symbolic recordings into genre categories. The best ranking results were presented by [3] with 77.17% and 65.28% mean hierarchical and raw classification accuracy, respectively.

In [3] a short account of a system that extracts 109 musical features from symbolic recordings and uses them to classify the recordings by genre is presented. The features used are based on instrumentation, texture, rhythm, dynamics, pitch statistics, melody and chords. The achieved reported classification reaches 90% for intra-category subcategories and 98% for categories. Though, this approach has an increased execution time while the space reduction is limited. The execution times (as seen from the results of the MIREX contest) are prohibitive for applications that require responses in real time, especially when frequent content update is potential. Additionally, the increased execution times were for a small database of 950 songs. Overall, the required methods need focus on a small selection of features which deliver increased selectivity performance. Moreover, the approach described in [3] additionally requires training for the"fittest" set of features.

Finally, Basili et al. [1] presented five features based on melody, timbre and rhythm for the purposes of symbolic music genre classification. Though, their investigation was oriented towards the comparison of different machine learning algorithms in genre classification.

### 2.1 Pitch Histograms

The authors of [5] introduced pitch histograms as a means to represent the pitch content of the notes of both symbolic and acoustic musical data. MIDI data files were used to extract note pitches, the frequency of occurrence of which constitutes the pitch histogram. As MIDI specification allows only for 128 discrete notes, each pitch histogram is an array of 128 values, indexed by the note number, representing the frequency of appearance of each note.

Tzanetakis et al. considered two versions of the pitch histogram according to whether the octave discrimination of notes is taken into consideration or not. Thus, the unfolded version does consider octaves in pitches of the notes leading to two C notes, being one octave apart, to be considered as, two different notes. In the folded version, all note pitches are transposed into a single octave, that is the two C notes of the previous example would be the same note, and then are mapped to a circle of fifths, so that adjacent histogram bins are spaced a fifth apart, rather than a semitone.

The rationale these choices rely on is that unfolded histograms can capture the pitch range of a piece, folding supports octave independency and the mapping to the circle of fifths ameliorates the expression of tonal music.

In order to minimise the search space, four one-dimensional features were extracted from the two histograms (folding and non-folding), namely PITCH-Fold, AMPL-Fold, PITCH-Unfold & DIST-Fold. The first is the bin number of the maximum peak of the folded histogram. The second is the amplitude of the maximum peak of the folded histogram. PITCH-Unfold is period of the maximum peak of the unfolded histogram, while DIST-Fold is the interval (in bins) between the two highest peaks of the folded histogram.

### 2.2 Motivation

Although the work of Tzanetakis et al. is rather intuitive, easy to perform, fast to calculate and the results reported are 1.6 times better than random classification, the accuracy still remains at levels that allow further amelioration. This is especially true, considering that each note carries additional information to pitch, its duration, that can equivalently easily be extracted and would not burden the dimensionality of the search space, at least to the point of recompensation by increasing the accuracy.

The use of the note duration is intuitively supported by the connection of note duration with rhythm. In a simplistic point of view, rhythm can be perceived as the number of notes within a bar, played at a specific tempo. As the total duration of notes within a bar is explicitly defined, smaller duration values lead to more notes within a bar, thus making the rhythm faster. The effect of the note durations on rhythm is rather important, as the music genres, generally, abide to rhythmic patterns.

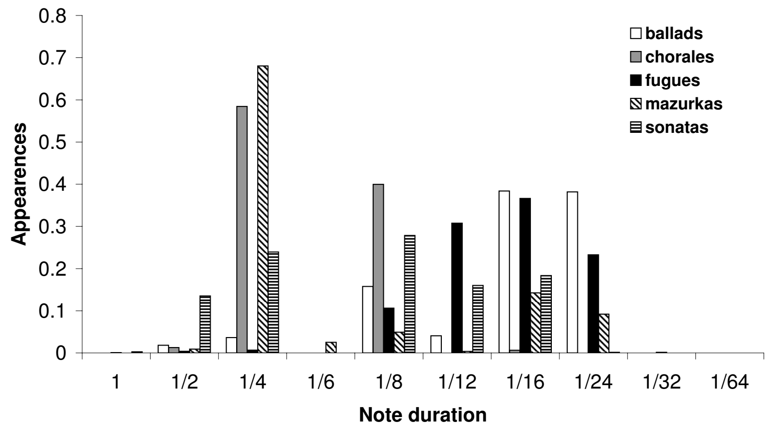
The aforementioned arguments broadly appear in Figure 1 where it can be seen that different musical genres have different frequencies of appearance for each note duration.

## 3 The Proposed Method

This work's key proposal is the utilisation of the duration histograms for the purposes of symbolic music genre classification.

This section presents three features that are based on the note duration dimension of a musical piece, as well as a differentiated, with respect to [5], approach in the extraction of features from the note pitch information of a piece.

A duration histogram is an array of 25 integer values (the eight standard durations, their dotted and double dotted augmentations and the breve duration) indexed by duration size that represents the frequency of occurrence of each note duration in a musical piece. Intuitively, duration histograms offer a means




(a)

**Fig. 1.** Normalised duration histogram

to capture the structure and rhythmic part of a piece. This is especially true in classical music where musical genres were created and evolved based on rules. For example, it is quite common for fugues to have several parts where the durations of the notes therein are significantly shorter than the other parts, in order to convey a sense of tenseness, since the original theme of fugues was an escape. On the other hand, sonatas are known to be structured to be more slow especially in their second parts.

As already discussed, the feature selection process is of great importance for all information retrieval purposes. This work proposes the extraction of three one-dimensional features from the duration histograms, namely the duration that has the greater frequency of appearance in a piece, the number of appearances of the duration with the highest frequency and the distance between the two highest frequency durations in terms of relative temporal duration.

The selection of the feature set is highly important, since the performance of the classifier mainly depends on the selectivity capability of the features to filter out statistical properties of the histogram that are irrelevant while retaining information that describe genre differences and thus assist the classification.

Accordingly, the proposed selection of features was based on the specific characteristics required to retain such as the note duration that appears more often as well as the second (indirectly through the distance) and the appearances of the most often duration. Additionally, features of the same style have successfully been employed in the literature for the purposes of symbolic genre classification, although on differentiated characteristic of the musical data.

### 3.1 The Proposed Features

**Non-folding Pitch** In non-folding pitch features, the effect of folding is not taken into consideration. Accordingly, the four one-dimensional feature vectors described in Section 2.1 are extracted based solely on unfolded histograms. This is done in order to establish the effect of folding in the classical works examined herein.

**Duration** In duration features, all features (as described in Section 3) are derived solely from duration histograms in order to determine the selectivity of the feature vectors produced by the duration histograms.

**Pitch & Duration** This feature is the combination of the feature vectors of the pitch information of the notes combined with the feature vectors produced by the duration information. Thus, each musical piece is represented by seven feature vectors, four from the pitch histogram and three from the duration histogram. As, pitch histogram can exist in two versions, the pitch & duration (or combination) feature vectors come in two flavours as well, the folding and non-folding.

**Weighted Pitch & Duration** The last feature proposed in this paper is a modified version of the combination feature previously described. The modification consists of a weighting scheme that allows the prediction of the genre to be more or less affected by one of the two features, in order to determine their contribution.

## 4 Performance Evaluation

In support of the efficiency of the proposed features, this section presents the experiments that have been performed. A concise description of the experimentation platform and data sets is also given followed by a performance analysis based on experimental comparison of the baseline and proposed features.

### 4.1 Experimental Set-up

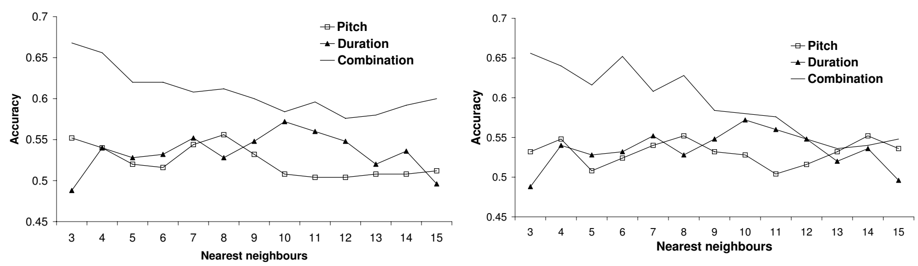
All algorithms described have been implemented and performed on a personal computer with 3,06GHz Intel Pentium IV processor, 1 GByte RAM, MS Windows XP operating system while the developing package utilised was MS Visual C++. The performance measure was the precision accuracy of the k-NN classifier.

The data sets employed for the experiments include real music objects, that originated from **kern Humdrum files acquired from the Humdrum website [6]. Each **kern file was stripped in order to retain only the note pitch and duration information. All the music objects pertain to classical works. The following five sub-categories were selected: ballads, chorales, fugues, mazurkas & sonatas and 50 songs were randomly selected by each category, adding up to a total corpus of 250 pieces.

After the vector extraction is completed (as described in Section 3), the distinguishing capability of the feature vectors is examined by means of the k-NN classifier, using the “leave one out” method. That is, one musical piece in the database is assumed to be of unknown genre and the rest of the pieces are considered as training data. Of the k nearest neighbour genres to the unknown piece, the genre with majority of appearances is predicted to be the genre of the piece assumed to be unknown. The process is repeated for all pieces in the database leading to the accuracy of the features.

### 4.2 Results

Initially, pitch, duration and combination were considered separately for both folding (pitch and combination) and non-folding features and the accuracy results are illustrated in Figure 2a and Figure 2b, respectively.




(a) (b)

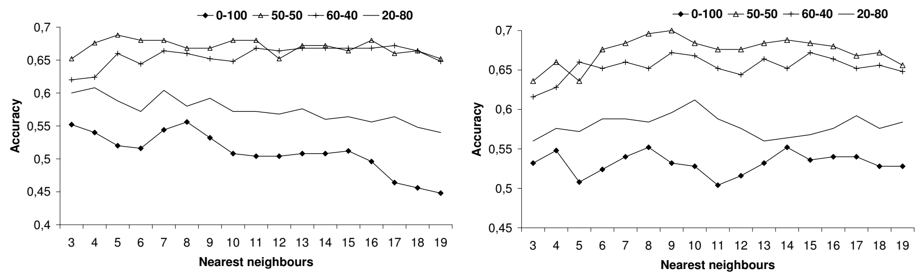
**Fig. 2.** Accuracy of all approaches (a) folding and (b) non-folding

Although pitch and duration have quite similar performances, duration is slightly better, while the combination features clearly outperform both. In addition, the changes in accuracy of the folding affected features seem to be rather marginal.

The next experimentation refers to the combination weighted approaches. Figure 3a and Figure 3b, provide four of the most characteristic weighting selections for both folding and non-folding. The legent titles imply the percentage of duration - pitch that participated.

Once again, the “0-100” combination, in which duration is not participating, presents the worse accuracy, while even 20% of duration contribution offers 9% increase in accuracy. The best performance is produced by the equally balanced participation of both duration and pitch (“50-50”). It should be noted that the performances between folding and non-folding are again quite similar. From this point on, folding features have not been considered further and all results imply non-folding features, where applicable.

Next, we experimented (Figure 4) on the size of the musical piece regarded as unknown (L). The size was adjusted by means of the number of notes within,




(a) (b)

**Fig. 3.** Accuracy of weighted combination (a) folding and (b) non-folding

while for the pitch features, cropping occurred at the ending of the piece, retaining, thus, the first L notes. For the pieces that had less a number of notes than L, the full piece was selected.

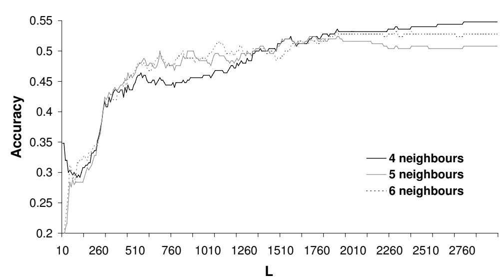



**Fig. 4.** Pitch accuracy for varying query size

An obvious trend towards better accuracy is clearly depicted in Figure 4, while for very small L the accuracy is equivalent to random (for five categories) since the note number is not enough for an accurate prediction.

In the following experiment, the size of the musical piece regarded as unknown (L) was examined against the accuracy for the duration features. In this experiment we additionally tested the offset the L notes are taken from. Figure 5 depicts the accuracy for L notes of the unknown genre piece taken from the middle part of the datum.

A slightly better performance in accuracy for the duration features is apparent in comparison to pitch features (Figure 4). The results for different offsets of the L notes proved identical making clear that the offset from which the part of the musical piece regarded of unknown genre is taken has no significant effect in the accuracy.

Following, is the experiment of the combination feature against the size of the musical piece regarded as unknown (Figure 6). Once again the performance

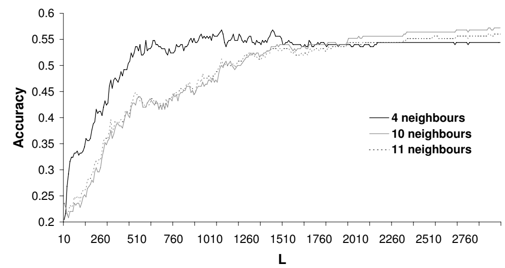



**Fig. 5.** Duration accuracy for varying query size taken from the middle of a file

of the combination approach is clearly better, in comparison to the accuracy results gathered for both pitch and duration separately.

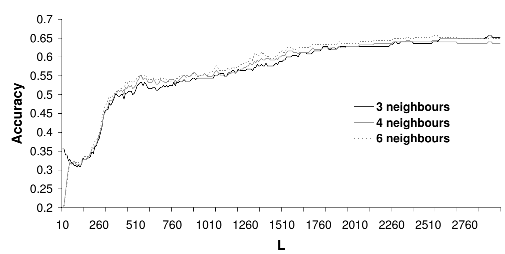
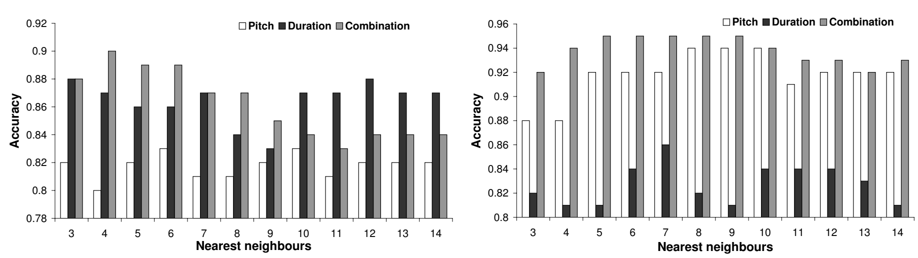



**Fig. 6.** Combination accuracy for varying query size

Finally, the last experiment performed a pairwise comparison between different genres. In this case, herein are presented two of the most representative results, the comparison between fugues - mazurkas (Figure 7a) & ballads - mazurkas (Figure 7b).




(a) (b)

**Fig. 7.** Accuracy for pairwise comparison between (a) fugues - mazurkas & (b) ballads - mazurkas

In Figure 7a, we observe the domination of the duration features over the pitch features, while in Figure 7b pitch features perform far better than duration features, though, the combination features are overall better.

## 5 Conclusions

This paper proposes the use of note pitch and duration histograms for the purposes of symbolic music genre classification. Note information histograms have a great capability in capturing a fair amount of information regarding harmonic as well as rhythmic features of different musical genres and pieces.

This paper proposes the incorporation of the note duration information during the feature extraction process. The duration dimension of a note is highly capable of supporting genre classification, though its weighted use with the pitch information proves even better.

This is verified through extensive experimental results, which illustrate the suitability of the proposed feature, reaching an accuracy level of 70%, that is a gain of 40% from the baseline approach.

Future work includes plans to examine broader ranges of musical categories, in order to establish the suitability of the proposed features, as well as the incorporation of the notion of patterns in genre classification. Patterns have played a significant role in the indexing of music even since the very first attempts of the creation of musical dictionaries.

## References


1. R. Basili, A. Serafini, and A. Stellato. Classification of musical genre: a machine learning approach. In *Proceedings of ISMIR*, 2004.
2. D. Byrd and T. Crawford. Problems of music information retrieval in the real world. *Information Processing and Management*, 38(2):249–272, 2002.
3. C. McKay and I. Fujinaga. Automatic genre classification using large high-level musical feature sets. In *Proceedings of ISMIR*, pages 31–38, 2004.
4. A. C. North and D. J. Hargreaves. Liking for musical styles. *Musicae Scientiae*, 1:109–128, 1997.
5. G. Tzanetakis, A. Ermolinskyi, and P. Cook. Pitch histograms in audio and symbolic music information retrieval. In *Proceedings of ISMIR*, pages 31–38, 2002.
6. The Humdrum website. A library of virtual musical scores in the humdrum **kern data format. http://kern.humdrum.net.