

# An HMM-Based Framework for Supporting Accurate Classification of Music Datasets

Alfredo Cuzzocrea<sup>1</sup>, Enzo Mumolo<sup>2</sup>, and Gianni Vercelli<sup>3</sup>

<sup>1</sup> DIA Dept., University of Trieste and ICAR-CNR, Italy  
alfredo.cuzzocrea@dia.units.it

<sup>2</sup> DIA Dept., University of Trieste, Italy  
mumolo@units.it

<sup>3</sup> DIBRIS Dept., University of Genova, Italy  
gianni.vercelli@unige.it

**Abstract.** In this paper, we use *Hidden Markov Models* (HMM) and *Mel-Frequency Cepstral Coefficients* (MFCC) to build statistical models of classical music composers directly from the music datasets. Several musical pieces are divided by instruments (*String, Piano, Chorus, Orchestra*), and, for each instrument, statistical models of the composers are computed. We selected 19 different composers spanning four centuries by using a total number of 400 musical pieces. Each musical piece is classified as belonging to a composer if the corresponding HMM gives the highest likelihood for that piece. We show that the so-developed models can be used to obtain useful information on the correlation between the composers. Moreover, by using the maximum likelihood approach, we also classified the instrumentation used by the same composer. Besides as an analysis tool, the described approach has been used as a *classifier*. This overall originates an HMM-based framework for supporting accurate classification of music datasets. On a dataset of *String Quartet* movements, we obtained an average composer classification accuracy of more than 96%. As regards instrumentation classification, we obtained an average classification of slightly less than 100% for *Piano, Orchestra* and *String Quartet*. In this paper, the most significant results coming from our experimental assessment and analysis are reported and discussed in detail.

## 1 Introduction

In the field of *music information analysis*, much research has been done in music genre classification, where a model is introduced in order to assign an unknown musical piece a score to a certain class, for example style, period, composer or region of origin. Two main categories of models have been used to do that: *global feature models* (e.g., [16]) and *n-gram models* (e.g., [29]). Global feature models represents every musical piece as a feature vector and use standard machine learning classifiers, whereas n-gram models rely on sequential features and use sequential machine learning algorithms, such as *Hidden Markov Models* (HMM) (e.g., [15]).

In this paper, we apply features derived from the *source-filter model* [12] and HMMs to analyze and classify pieces of classical music, thus combining the two paradigms. This overall originates an HMM-based framework for supporting accurate classification of music datasets. Basically, in this work we adopt the approach that is usually applied to human voice, namely *Mel-Frequency Cepstral Coefficients* (MFCC) (e.g., [12]), and left-to-right HMMs. Indeed, these approaches, in recent initiatives (e.g., [1, 14, 4, 11, 27, 21, 33, 22]), have also been used for musical genre and instrument classification.

We have considered several types of music recording, both choral and instrumental, and we used them to develop statistical models using a classical sequential algorithm. The models have been used to find correlations between composers using the same instrument. Our results shows that this approach can be used both for composer analysis and composer classification. Moreover, the same approach has been used to classify the instrumentation used by a composer, but with some limitation: drums and similar instruments are not well described with the source-filter features and this is verified by our results. For other types of instruments, this method works in a satisfactory way.

Our experimental analysis developed in our research works as follows. We have divided the 400 musical works for composer and instruments. For example, considering *Mozart*, we selected ten pieces of *Piano* (in the following called *Mozart-Piano dataset*), ten pieces of *Orchestra* and ten pieces of *String Quartet* compositions of this author. Of the ten pieces, five pieces are used for the training set and five for the testing set. In addition to this, each piece is then split in two sub-pieces, as follows: only the initial 4 minutes of each piece are used for training the models and the following 4 minutes for testing purposes (of course, the pieces can be longer). By incremental training, we developed an HMM model of *Mozart-Piano*, a model of *Mozart-Orchestra* and a model of *Mozart-String Quartet*, and so forth for each composer. Then, all the available *Piano* compositions are used to compute the likelihood that the musical piece was produced by the *Mozart-Piano* HMM, and so forth for the other compositions. The *confusion matrix* obtained on a maximum likelihood basis describes the relations between the composers using the same instrument. As we show in our experimental assessment and analysis, the obtained results are well-justified on the basis of the known influence among composers. Therefore, a key contribution of this work is to show that left-right HMM and MFCC features can be used to obtain an accurate model of each composer, thus achieving a composite classification framework for music datasets. Indeed, the obtained statistical models can be used to describe the relations among the composers. Our experimental assessment and analysis confirms the benefits coming from our proposal.

The remaining part of this paper is organized as follows. In Section 2, previous work on classical music classification are reported and discussed. In Section 3, our proposed classification framework is described. In Section 4, the experimental results related to the classification of composers and instrumentation are reported. Final remarks and future work analysis conclude the paper in Section 5.

## 2 Related Work

Genre classification has been studied by Yaslan and Cataltepe [32]. They introduce a semi-supervised random feature ensemble method for audio classification that uses labeled and unlabeled data together. Genre classification of classical music has been addressed by Pollastri and Simoncelli [24] by using melody as features and HMM as classifier. They tested the algorithm using four composers. In [17], the MFCCs have been compared to other features for music classification, showing that they are simpler but effective for this purpose. Moreover, approaches based on MFCCs have been used to perform music segmentation (e.g., [3]). In [23], the detection of voice segments in music songs is described. This solution extracts the MFCCs of the sound and uses an HMM to infer if the sound has voice. Marinescu and Ramirez report in [20] an approach that aims at developing models of singers and use them to generate expressive performances similar in voice quality and style with the original singers. Their approach is based on applying machine learning to discover singer-specific timing patterns of expressive singing based on existing performances.

[25] proposes to classify latin music data by exploiting the “cifras” of the songs. [13] focuses on the classification of music data based on lyrics by using non-emotional words. Canonical correlation analysis is instead exploited in [2] to classify Greek folk music. Acoustic and visual features to support music data classification are discussed in [30], whereas global feature models and variable neighborhood search are exploited in [16].

Interesting are also some recent music classification approaches in the context of *deep learning*. For instance, [18] proposes to use deep neural networks to support music genre classification. Convolutional neural networks are instead argued in [34] to be a good classifier for music genre data.

Finally, supervised and semi-supervised methods are studied in [26] and [31], respectively.

## 3 A Framework for Supporting Accurate Classification of Music Datasets based on HMM

In this Section, we describe our proposed framework for supporting accurate classification of music datasets, by also highlighting the different types of models that we compare in our experimental campaign. Figure 1 provides an overview of our HMM-based classification framework. It essentially comprises two main components: (i) *Feature Extraction* and (ii) *DHMM*. The first one is the component devoted to extract suitable features from the target music dataset. The second one implements the *Discrete HMM* (DHMM) (e.g., [15]) embedded in our proposal. As we describe in Section 3.3, our choice was to select the discrete case as to take advantages from the discrete representation and mapping in data processing.

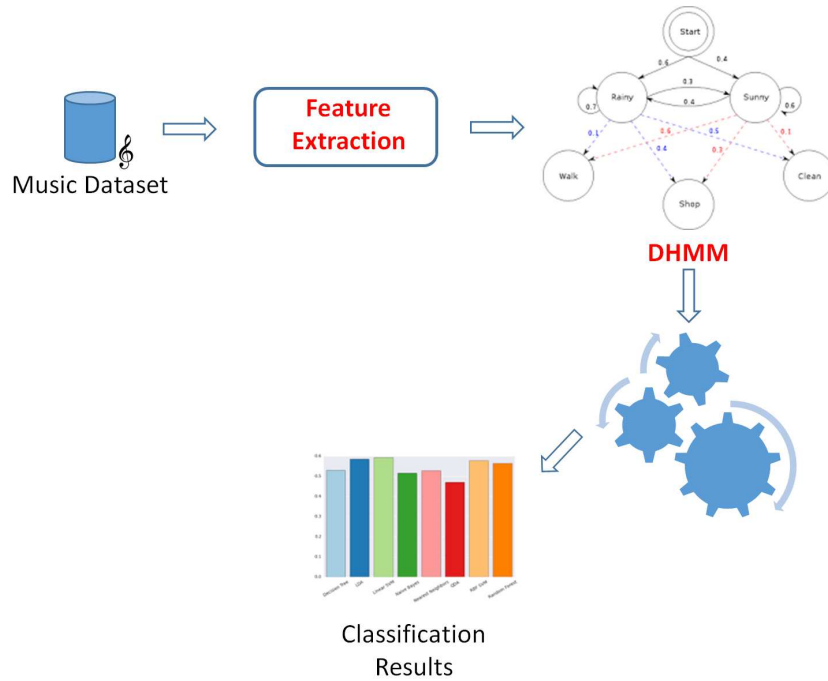


Fig. 1. HMM-Based Classification Framework

### 3.1 Music Dataset

Our target music dataset comprises a total number of 19 composers, with different styles spanning four centuries, from the 17th to the 20th century, namely: *Bach, Bartok, Beethoven, Berg, Chopin, Debussy, Dvorak, Haydn, Liszt, Mahler, Martinu, Mendelssohn, Mozart, Palestrina, Rihm, Schmittke, Schoenberg, Schumann, Stravinsky*. For the sake of completeness, some musical compositions used in this study are listed, by composer, in the following:

- *Bach*: choral works from BWV225 to BWV229, and the orchestral music from 1046 to 1051;
- *Beethoven*: piano sonatas from 7 to 57, string quartets from Op.127 to Op.135, orchestra symphonies from n.1 to n.9;
- *Chopin*: piano concerto from Op.11 to Op.21, piano suites from Op.22 to Op.52;
- *Mahler*: symphonies from n.1 to n.6;
- *Mozart*: string quartets from KV169 to KV465, orchestra symphonies from n.1 to n.6, piano sonatas from n.10 to n.14;
- *Debussy*: the piano compositions *Arabesque, Ballade, Pagodes, Danse, Images Hommage, Images Reflects, La Boite per Tableau, Nocturne, Preludio, La Cathedrale, Clair de Lune*.

### 3.2 Feature Extraction

A well-known, simple model of many sounds, like speech and some music, is the source-filter model [12], which we used as fundamental feature extraction model. According to this model, sound is the convolution between the impulse response of a filter with a source signal. For example, in wind instruments, the filter describe the acoustic characteristics of the instrument and the source signal is the signal produced by the reed. The problem is therefore to separate the filter from the source component, which requires ad-hoc deconvolution operations.

One of the approaches to solve this problem is to use *homomorphic deconvolution* [28]. Let  $x(n)$  the result of a convolution operation between two signals, the *cepstrum* of  $x(n)$  is obtained by taking the logarithm of the spectrum of  $x(n)$ . Under some circumstances, it is possible to separate the two components. To this end, we use a filter bank, spaced uniformly on a Mel scale, and we considered the output power of these filters in the center of each band-pass filter. We can interpret the output of a single band-pass filter as the  $k$ -th component of the DFT of the input sequence  $x(n)$ , denoted by  $X(k)$  and defined as follows:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{j\frac{2\pi}{N}nk} \quad (1)$$

Since we are interested in the center frequencies of the band-pass filters, the  $k$ -th frequency is moved by  $\frac{\pi}{N}$ . Therefore:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{j[\frac{2\pi}{N}k + \frac{\pi}{N}]n} = \sum_{n=0}^{N-1} x(n)e^{j\frac{2\pi}{N}n(k-0.5)} \quad (2)$$

On the basis of this approach, we compute the MFCCs as the  $x(n)$  obtained in terms of the inverse filter of Equation (2), as follows:

$$x(n) = \sum_{k=0}^M X(k) \cos(2\pi(k-0.5)n/N) \quad (3)$$

where  $X(k)$  is the logarithm of the output energy for the  $k$ -th triangular filter,  $N$  is the number of filters and  $M$  is the number of MFCCs.

### 3.3 Discrete Hidden Markov Models

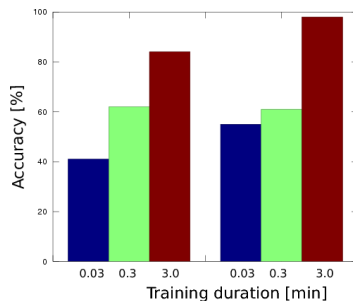
In our framework, we make use of DHMM. Markov models describe time series from a stochastic point of view, taking into account the correlations of signals (e.g., [15]).

In HMMs, the output for each state corresponds to an output probability distribution instead of a deterministic event. That is, if the observations are sequences of discrete symbols chosen from a finite alphabet, then, for each state, a corresponding discrete probability distribution describes the stochastic process to be modeled. It is worth noting that, in our framework, we use *discrete observations*, by converting the MFCCs in discrete symbols using standard vector

quantization [19], thus achieving the DHMM. The usage of DHMM is motivated by the fact that introducing less symbols makes the model simpler and, at the same, more solid. On the other hand, while it is still possible to convert the continuous observations into discrete ones using vector quantization, some performance degradation due to the quantization process is, however, introduced. Therefore, from a performance point-of-view, it is important to use an overall continuous formulation of algorithms, and consequently introduce ad-hoc optimizations. Our framework adheres to this computational paradigm.

## 4 Experimental Assessment and Analysis

In this Section, we provide the results of our experimental assessment and analysis devoted to stress the accuracy of the proposed classification framework. In our experimental campaign, we acquired all the sounds at 11025 Hz with 16 bits. The signal is divided in non overlapping frames of 200 samples, or 18.14 ms, for the subsequent processing. The first problem we explore in this study is the following: what is the best training duration of the DHMM? The second problem is the choice of the number of centroid of the DHMM. The third problem is the choice of the number of states of the DHMM. The last two problems have the following answer: the best results were obtained with 256 centroid and 48 states, so we used these parameters for all the experiments. Regarding the training duration, we computed the average classification accuracy over all the composers and all the instrumentation.



**Fig. 2.** Average Classification Accuracy for all the Composers (left group of 3 bars) and Instrumentation (right group of 3 bars), for Different Training Durations

In more details, Figure 2 reports, by using a bar shape, the average accuracies obtained for the composers and instrumentation, respectively, for three different DHMM training lengths over the 4 minutes basis (see Section 1): 0.03, 0.3 and 3.0 minutes. We observe that the minimum training length to get good enough results is 3.0 minutes. Furthermore, with more than 3.0 minutes DHMM training, the accuracy tends to be asymptotic. While this experiment provides average

results, it is more interesting to look at the detailed results given as confusion matrixes, as reported in the following Sections.

#### 4.1 Experimental Results for Composer Classification

In this Section, we report the classification results of composers using the same instrumentation, by using the confusion matrix model. The first test is concerned with the choral compositions of *Bach*, *Palestrina* and *Schnittke* (see Table 1). Looking at the confusion matrix, we observe that *Schnittke* is often confused with *Palestrina*. This is probably due to the fact that the Russian composers of the 20th century were influenced by the composition techniques of the Italian and Flemish schools of the 16th century.

<i>Chorus</i>	<i>Bach</i>	<i>Palestrina</i>	<i>Schnittke</i>
<i>Bach</i>	83.3 %	16.7 %	0 %
<i>Palestrina</i>	0 %	98.9 %	0 %
<i>Schnittke</i>	0 %	25 %	75 %

**Table 1.** *Chorus* Classification Performance

Then, we study the *String Quartet* compositions (see Table 2).

<i>Quartet</i>	<i>Bartok</i>	<i>Beethoven</i>	<i>Dvorak</i>	<i>Haydn</i>	<i>Martinu</i>	<i>Mendelssohn</i>	<i>Mozart</i>	<i>Rihm</i>	<i>Schoenberg</i>
<i>Bartok</i>	98.7 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
<i>Beethoven</i>	0 %	92 %	0 %	0 %	0 %	0 %	8 %	0 %	0 %
<i>Dvorak</i>	0 %	0 %	98.5 %	0 %	0 %	0 %	0 %	0 %	0 %
<i>Haydn</i>	0 %	0 %	0 %	98.6 %	0 %	0 %	0 %	0 %	0 %
<i>Martinu</i>	0 %	0 %	0 %	0 %	99.1 %	0 %	0 %	0 %	0 %
<i>Mendelssohn</i>	0 %	0 %	0 %	0 %	0 %	98.9 %	0 %	0 %	0 %
<i>Mozart</i>	0 %	0 %	0 %	0 %	0 %	0 %	99.2 %	0 %	0 %
<i>Rihm</i>	0 %	0 %	0 %	0 %	0 %	0 %	0 %	92 %	8 %
<i>Schoenberg</i>	0 %	0 %	0 %	0 %	0 %	0 %	0 %	16 %	84 %

**Table 2.** *String Quartet* Classification Performance

Looking at the confusion matrix in Table 2, we observe that some confusion between *Beethoven* and *Mozart*, and *Rihm* and *Schoenberg*, respectively, arise. The confusion between *Beethoven* and *Mozart* can be justified by the fact that

*Beethoven* was part of the transition of the Classical to the Romantic periods while *Mozart* was a composer of the Classical period. Moreover, we observe some confusion between *Rihm* and *Schoenberg*. Also, this results can be justified by the fact that *Rihm* is a German composer while *Schoenberg* was an Austrian composer, associated with the German expressionist movement, and the former work of *Rihm* was inspired by the *Schoenberg*'s expressionist period.

Next, we make some tests with the *Orchestra* compositions (see Table 3).

<i>Orchestra</i>	<i>Bach</i>	<i>Beethoven</i>	<i>Berg</i>	<i>Chopin</i>	<i>Mahler</i>	<i>Stravinsky</i>
<i>Bach</i>	33.3 %	0 %	33.3 %	33.3 %	0 %	0 %
<i>Beethoven</i>	0 %	83.3 %	16.7 %	0 %	0 %	0 %
<i>Berg</i>	0 %	0 %	83.3 %	16.7 %	0 %	0 %
<i>Chopin</i>	0 %	33.3 %	0 %	66.7 %	0 %	0 %
<i>Mahler</i>	0 %	0 %	50 %	16.7 %	33.3 %	0 %
<i>Stravinsky</i>	16.6 %	0 %	16.6 %	50 %	16.6 %	0 %

**Table 3.** *Orchestra* Classification Performance

From the confusion matrix in Table 3, there are some results to be discussed. The most notably result is probably that *Stravinsky* failed to be recognized as himself, and this can be justified by the fact that he used a lot of percussion in his compositions, and the MFCC, which are derived by a source-filter model of the signal being analyzed, do not describe well this type of sound. Another notably result is that *Stravinsky*'s models are confused with *Chopin*. This result could be justified by the fact that that the musical pieces used for training *Chopin*'s models for orchestral music were concerts for *Piano* and *Orchestra* (mainly Op.11) that inspired *Stravinsky*'s work. On the other hand, *Chopin* is not confused with *Stravinsky* simply because he died before *Stravinsky*. There is also some confusion between *Bach* and *Chopin* while it is not true the opposite. In fact, *Bach*'s model were obtained using *Bach*'s Brandenburg Concertos which did not use instruments typical of *Chopin*'s orchestral music. We also observe an important confusion between *Mahler*'s models and *Berg*'s models, while the opposite does not hold. The linkage between *Mahler* and *Berg* can be explained by the fact that *Mahler* was related with *Schoenberg* as regards for example the instrumental technique, and *Berg* was a *Schoenberg* alumnus.

The classification performances are generally good, with the exception of *Chopin*, *Bach* and *Stravinsky*. Concerning the orchestral compositions of *Chopin*, it is worth noting that his music considered for training the DHMM are concerts for piano and orchestra. Concerning *Bach*, his Brandenburg Concertos didn't use percussion or other instruments typical of the orchestral music of following years, as *Chopin* used. The *Stravinsky*'s poor results can be justified by the fact that he use a lot of percussion, and the source-filter model is not able to model this type of sound.



The last experiments are concerned with the compositions for *Piano* (see Table 4).

<i>Piano</i>	<i>Beethoven</i>	<i>Chopin</i>	<i>Debussy</i>	<i>Liszt</i>	<i>Mozart</i>	<i>Schumann</i>
<i>Beethoven</i>	98.7 %	0 %	0 %	0 %	0 %	0 %
<i>Chopin</i>	0 %	66.6 %	0 %	0 %	16.6 %	16.6 %
<i>Debussy</i>	0 %	16.7 %	50 %	0 %	0 %	33.3 %
<i>Liszt</i>	0 %	33.3 %	0 %	33.3 %	16.6 %	16.6 %
<i>Mozart</i>	0 %	0 %	0 %	0 %	66.7 %	33.3 %
<i>Schumann</i>	50 %	0 %	0 %	0 %	0 %	50 %

**Table 4.** *Piano* Classification Performance

In this case, an important result is the confusion between *Debussy* and *Schumann*. This may mean that *Debussy* was greatly inspired by *Schumann*. We note also that in some cases *Debussy* is confused with *Chopin*. In fact, *Debussy* was a *Chopin*'s alumnus. Also, there is an important confusion between *Schumann* and *Beethoven*, which may mean that the *Beethoven*'s influence on *Schumann* was quite important. From the results of Table 4, it is worth noting that there is an important confusion between *Liszt* and *Chopin*. In fact, under the influence related with the friendship with *Chopin*, *Liszt*'s developed his poetic and romantic side.

A first conclusion of this set of experiments is that the features derived from the source-filter model are best suitable to model speech (coral compositions), *String* and *Piano* sounds. A second conclusion is that the classification results are well justified in terms of the influence of one composer on another one, thus meaning that the described method can be used to study the influence between composers.

## 4.2 Experimental Results for Instrumentation Classification

In this Section, we report the classification results of the instrumentation by the same composer, by using the confusion matrix model. In this case we also considered three different instruments: *Piano Solo*, *String Quartet* and *Orchestra*. In Table 5 and Table 6, we observe that these three settings are successfully recognized for *Beethoven*'s and *Mozart*'s works.

We have also obtained good classification of instrumentation for *Haydn*'s (see Table 7) and *Schubert*'s (see Table 8) works.

Even from the second set of experiments, we can conclude that our proposed framework provides a good accuracy in the case of instrument classification too, thus proving its flexibility and reliability.

<i>Beethoven</i>	<i>Piano</i>	<i>Orchestra</i>	<i>Quartet</i>
<i>Piano</i>	98.4 %	0 %	0 %
<i>Orchestra</i>	0 %	98.7 %	0 %
<i>Quartet</i>	0 %	0 %	98.9 %

**Table 5.** Instrument Classification of *Beethoven*'s Works

<i>Mozart</i>	<i>Piano</i>	<i>Orchestra</i>	<i>Quartet</i>
<i>Piano</i>	99.1 %	0 %	0 %
<i>Orchestra</i>	0 %	98.2 %	0 %
<i>Quartet</i>	0 %	0 %	98.8 %

**Table 6.** Instrument Classification of *Mozart*'s Works

<i>Haydn</i>	<i>Piano</i>	<i>Orchestra</i>	<i>Quartet</i>
<i>Piano</i>	98.3 %	0 %	0 %
<i>Orchestra</i>	10 %	90 %	0 %
<i>Quartet</i>	0 %	0 %	98.8 %

**Table 7.** Instrument Classification of *Haydn*'s Works

<i>Schubert</i>	<i>Piano</i>	<i>Orchestra</i>	<i>Quartet</i>
<i>Piano</i>	99.3 %	0 %	0 %
<i>Orchestra</i>	0 %	90 %	10 %
<i>Quartet</i>	0 %	0 %	99.7 %

**Table 8.** Instrument Classification of *Schubert*'s Works

## 5 Final Remarks and Future Work

In this paper, we have proposed a classification framework that exploits the source-filter approach and HMM machine learning techniques for classifying musical recordings. While our framework is rather simple, it proved to ensure good classification accuracy. The experimental results show that these methodologies are well-suited for modeling classical music even when no voices are present in the recordings. A major limitation of this method is due to the weak behaviors in modeling drums, as the feature extraction technique we used removes quite completely such frequencies. This can lead to a modification of the approach to take into account this aspect in future works. Another important line of research to explore concerns with making the proposed framework able to deal with novel challenges posed by emerging *big data trends* (e.g., [9, 10, 7, 5, 6, 8]).

## References

1. J. Ajmera, I. McCowan, and H. Bourlard. Speech/music segmentation using entropy and dynamism features in a HMM classification framework. *Speech Communication*, 40(3):351–363, 2003.
2. N. Bassiou, C. Kotropoulos, and A. Papazoglou-Chalikias. Greek folk music classification into two genres using lyrics and audio via canonical correlation analysis. In

- 9th International Symposium on Image and Signal Processing and Analysis, ISPA 2015, Zagreb, Croatia, September 7-9, 2015*, pages 238–243, 2015.
3. D. G. Bhalke, C. B. R. Rao, and D. S. Bormane. Automatic musical instrument classification using fractional fourier transform based- MFCC features and counter propagation neural network. *J. Intell. Inf. Syst.*, 46(3):425–446, 2016.
  4. A. Cont. Realtime audio to score alignment for polyphonic music instruments, using sparse non-negative constraints and hierarchical HMMS. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing, ICASSP 2006, Toulouse, France, May 14-19, 2006*, pages 245–248, 2006.
  5. A. Cuzzocrea. Accuracy control in compressed multidimensional data cubes for quality of answer-based OLAP tools. In *18th International Conference on Scientific and Statistical Database Management, SSDBM 2006, 3-5 July 2006, Vienna, Austria, Proceedings*, pages 301–310, 2006.
  6. A. Cuzzocrea. Privacy and security of big data: Current challenges and future research perspectives. In *Proceedings of the First International Workshop on Privacy and Security of Big Data, PSBD@CIKM 2014, Shanghai, China, November 7, 2014*, pages 45–47, 2014.
  7. A. Cuzzocrea, F. Furfaro, and D. Saccà. Enabling OLAP in mobile environments via intelligent data cube compression techniques. *J. Intell. Inf. Syst.*, 33(2):95–143, 2009.
  8. A. Cuzzocrea and U. Matrangolo. Analytical synopses for approximate query answering in OLAP environments. In *Database and Expert Systems Applications, 15th International Conference, DEXA 2004 Zaragoza, Spain, August 30-September 3, 2004, Proceedings*, pages 359–370, 2004.
  9. A. Cuzzocrea, D. Saccà, and J. D. Ullman. Big data: a research agenda. In *17th International Database Engineering & Applications Symposium, IDEAS '13, Barcelona, Spain - October 09 - 11, 2013*, pages 198–203, 2013.
  10. A. Cuzzocrea, I. Song, and K. C. Davis. Analytics over large-scale multidimensional data: the big data revolution! In *DOLAP 2011, ACM 14th International Workshop on Data Warehousing and OLAP, Glasgow, United Kingdom, October 28, 2011, Proceedings*, pages 101–104, 2011.
  11. V. Emiya, R. Badeau, and B. David. Automatic transcription of piano music based on HMM tracking of jointly-estimated pitches. In *2008 16th European Signal Processing Conference, EUSIPCO 2008, Lausanne, Switzerland, August 25-29, 2008*, pages 1–5, 2008.
  12. G. Fant. *Acoustic Theory of Speech Production*. Mouton, The Hague, 1960.
  13. M. Furuya, K. Oku, and K. Kawagoe. Music feeling classification based on lyrics using weighting of non-emotional words. In *Proceedings of the 13th International Conference on Advances in Mobile Computing and Multimedia, MoMM 2015, Brussels, Belgium, December 11-13, 2015*, pages 380–383, 2015.
  14. S. Gao and Y. Zhu. A hmm-embedded unsupervised learning to musical event detection. In *Proceedings of the 2005 IEEE International Conference on Multimedia and Expo, ICME 2005, July 6-9, 2005, Amsterdam, The Netherlands*, pages 334–337, 2005.
  15. Z. Ghahramani. An introduction to hidden markov models and bayesian networks. *IJPRAI*, 15(1):9–42, 2001.
  16. D. Herremans, K. Sørensen, and D. Martens. Classification and generation of composer-specific music using global feature models and variable neighborhood search. *Computer Music Journal*, 39(3):71–91, 2015.

17. J. H. Jensen, M. G. Christensen, M. Murthi, and S. H. Jensen. Evaluation of mfcc estimation techniques for music similarity. In *European Signal Processing Conference, EUSIPCO*, 2006.
18. I. Jeong and K. Lee. Learning temporal features using a deep neural network and its application to music genre classification. In *Proceedings of the 17th International Society for Music Information Retrieval Conference, ISMIR 2016, New York City, United States, August 7-11, 2016*, pages 434–440, 2016.
19. Y. Linde, A. Buzo, and R. M. Gray. An algorithm for vector quantizer design”. *IEEE Transactions on Communications*, pages 702–710, 1980.
20. M.-C. Marinescu and R. Ramirez. Modeling expressive performances of the singing voice. In *International Workshop on Machine Learning and Music*, 2009.
21. J. Myung, K. Kim, J. Park, M. Koo, and J. Kim. Two-pass search strategy using accumulated band energy histogram for hmm-based identification of perceptually identical music. *Int. J. Imaging Systems and Technology*, 23(2):127–132, 2013.
22. E. Nakamura, K. Yoshii, and S. Sagayama. Rhythm transcription of polyphonic piano music based on merged-output HMM for multiple voices. *IEEE/ACM Trans. Audio, Speech & Language Processing*, 25(4):794–806, 2017.
23. R. Nobrega and S. Cavaco. Detecting key features in popular music: case study - singing voice detection. In *International Workshop on Machine Learning and Music*, 2009.
24. E. Pollastri and G. Simoncelli. Classification of melodies by composer with hidden markov models. *Web Delivering of Music, International Conference on*, 0:0088, 2001.
25. A. L. Przybysz, R. Corassa, C. L. dos Santos, and C. N. S. Jr. Latin music mood classification using cifras. In *2015 IEEE International Conference on Systems, Man, and Cybernetics, Kowloon Tong, Hong Kong, October 9-12, 2015*, pages 1682–1686, 2015.
26. B. Rajesh and D. G. Bhalke. Automatic genre classification of indian tamil and western music using fractional MFCC. *I. J. Speech Technology*, 19(3):551–563, 2016.
27. J. C. Ross and J. Samuel. Hierarchical clustering of music database based on HMM and markov chain for search efficiency. In *Speech, Sound and Music Processing: Embracing Research in India - 8th International Symposium, CMMR 2011, 20th International Symposium, FRSM 2011, Bhubaneswar, India, March 9-12, 2011, Revised Selected Papers*, pages 98–103, 2011.
28. J. M. Tribolet. *Seismic applications of homomorphic signal processing*. Prentice Hall, 1979.
29. J. Wolkowicz and V. Keselj. Evaluation of n-gram-based classification approaches on classical music corpora. In *Mathematics and Computation in Music - 4th International Conference, MCM 2013, Montreal, QC, Canada, June 12-14, 2013. Proceedings*, pages 213–225, 2013.
30. M. Wu and J. R. Jang. Combining acoustic and multilevel visual features for music genre classification. *TOMCCAP*, 12(1):10:1–10:17, 2015.
31. X. Yang, L. He, D. Qu, W. Zhang, and M. T. Johnson. Semi-supervised feature selection for audio classification based on constraint compensated laplacian score. *EURASIP J. Audio, Speech and Music Processing*, 2016:9, 2016.
32. Y. Yaslan and Z. Cataltepe. Audio genre classification with semi-supervised feature ensemble learning. In *International Workshop on Machine Learning and Music*, 2009.

33. Y. R. Yeminy, Y. Keller, and S. Gannot. Single microphone speech separation by diffusion-based HMM estimation. *EURASIP J. Audio, Speech and Music Processing*, 2016:16, 2016.
34. W. Zhang, W. Lei, X. Xu, and X. Xing. Improved music genre classification with convolutional neural networks. In *Interspeech 2016, 17th Annual Conference of the International Speech Communication Association, San Francisco, CA, USA, September 8-12, 2016*, pages 3304–3308, 2016.