Novel use of the variogram for MFCCs modeling

Simone Sammartino, Lorenzo J. Tardón, and Isabel Barbancho

Dept. Ingeniería de Comunicaciones, E.T.S. Ingeniería de Telecomunicación,
Universidad de Málaga, Campus Universitario de Teatinos s/n, 29071, Málaga, Spain
{lorenzo,ibp}@ic.uma.es

Abstract. The paper describes two novel variants of the use of the variogram as summarizing tool for the MFCCs. A full variogram calculated on the second MFCC and a reduced variogram calculated on a subset of distance lags on the whole MFCCs matrix (first coefficient excluded), are proposed as tools to synthesize the timbre information of the MFCCs, for music similarity. Also, four different weighting functions are tested for the calculus of the (Euclidean) distance among the songs. The performance of the methods is evaluated by the application of the pseudo-objective evaluation of the MIREX AMS task, and compared with the scores of the methods submitted to the MIREX AMS 2011.

Keywords: Music similarity, MFCCs, variogram, MIREX AMS, musical genre classification

1 Introduction

The massive improvement of the Internet communication technology over the last years allowed the fast development of on-line games, multimedia playing and digital content sharing. The advances in the distribution of music contents led to the urgent need of a proper storage, labelling and indexation of the material, with the aim of an efficient access and retrieving of the items. One of the most demanded task is the automatic recommendation of music contents, aimed to help the user to choose a track with the highest degree of similarity with some defined references.

One of the fields where the MIR community is currently investing more resources are the so called *content-based* music recommendation systems where music similarity is evaluated on the basis of the calculus of a number of descriptors from time and frequency domain, and the derivation of some kind of feature patterns that are used as signature of the songs.

One of the most successfully used features to describe the spectral content of an audio signal are the Mel Frequency Cepstral Coefficients (MFCCs) [14]. These short-term spectral-based features are popularly employed to summarize the timbre content of the song and they are involved in most of the known algorithms for music similarity.

The MFCCs are calculated according to a recognized standard procedure: 1) calculus of the short-term spectrogram, 2) mapping of the spectrogram on

9th International Symposium on Computer Music Modelling and Retrieval (CMMR 2012)
19-22 June 2012, Queen Mary University of London

the Mel scale, through the application of a Mel frequencies bank filter, 3) transformation of the filtered spectrum to decibels and finally 4) compression of the resulting matrix by the application of the Discrete Cosine Transform.

Due to the own scheme of calculation of the MFCCs, the resulting descriptor is a matrix whose size depends both on the number of coefficients (fixed a priori) and the set of chunks in which the song has been fractioned during the windowing of the spectrogram. For this reason, the use of the MFCCs is usually associated with some kind of clustering of the coefficients, in order to represent the timbre descriptor as a fixed-size compressed matrix, to be employed directly as a standardized signature for the audio signals.

Logan and Salomon [10], employed the popular K-means method to cluster the MFCCs and used the means and covariance matrices of the centroids to define the song signature. Pampalk [13] proposed the use of the Gaussian Mixture Models (GMM) and the Expectation-Maximization (EM) approach, by modelling the probability distribution functions of the coefficients vectors. Aucouturier and Pachet [1] employed the Monte Carlo approach as clustering technique, Mandel and Ellis [11] used only one cluster from GMM, while Tzanetakis and Cook [17] simply extracted the mean and variance from each vector of Mel coefficients. In [15], Sammartino et al proposed the use of the variogram for MFCCs modelling. In this work, two novel variants of the calculus of the variogram are analyzed and their performance is evaluated with different setups.

The article is organized as follows: after this brief introduction on the music similarity framework, the use of the variogram as summarizing tool for the MFCCs is detailed in Section 2. The results of the evaluation of the methods proposed are presented in Section 3 and finally, some conclusions are drawn in Section 4.

2 The variogram

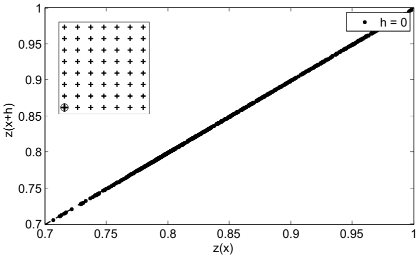
The variogram is a very popular tool in Geostatistics, widely employed to model the spatial continuity of environmental variables. Isaaks and Srivastava [5] affirm that "*Two data close to each other are more likely to have similar values than two data that are far apart.*" This characteristic is quantitatively defined as *spatial continuity*, referring to the spatial correlation of spatial variables.

2.1 The spatial variogram

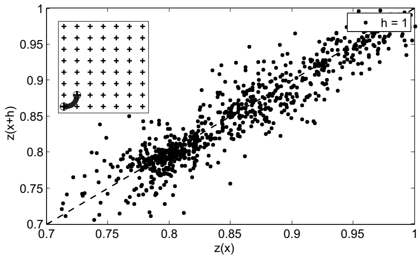
Let $z(x)$, with $x = 1, \dots, n$ represents a set of n (regularly) sampled observations of a spatial phenomenon. The term x stands for the vector of spatial bi- or three-dimensional coordinates of the samples (generally unidimensional in the case of temporal variables).

One way to measure the spatial continuity of the samples is to observe how they behave when paired by their reciprocal distance. The h-scatter plot fulfills this target. It is the scatter plot of samples paired by a specific value of distance h [5].

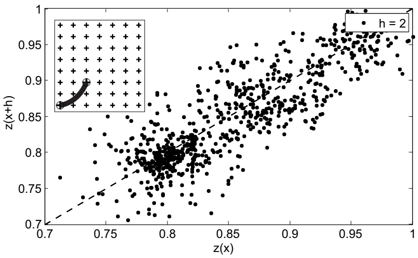
In Figure 1, three examples of h-scatter plots are shown. The samples are paired by a three distance vectors of $h = \{1, 2, 5\}$ (more precisely, it is $h = \{1, 2, 5\}$ in both x and y axis), and represented as points scattered over the bisector. Additionally, the h-scatter plot of the points coupled with themselves ($h = 0$) is presented (Fig. 1(a)). The bisector represents the geometrical locus of the pairs of samples separated by zero distance (all the samples paired with themselves).



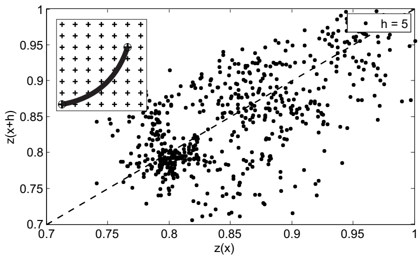
(a) The h-scatter plot calculated with samples paired with themselves ($h = 0$).



(b) The h-scatter plot calculated with samples paired at a distance $h = 1$.



(c) The h-scatter plot calculated with samples paired at a distance $h = 2$.



(d) The h-scatter plot calculated with samples paired at a distance $h = 5$.

Fig. 1. h-scatter plots of a set of regularly gridded spatial samples. The first plot is perfectly aligned with the diagonal. In fact the act of pairing samples at distance $h = 0$ means comparing each samples with itself. In the other plots the increase of the spread of the cloud is evident. In each plot, a small graph showing the pairing of the first sample is shown. The analysed data are a subset of the topographic data, provided by the US National Geophysical Data Centre (NOAA).

Following the definition of the spatial continuity, we can model how the spread of the clouds of points varies with the distance h. The greater the distance of the paired points, the fatter the cloud and the larger the difference between the samples of each pair. Therefore, it can be assumed that the spread of the cloud will range between zero, when the distance is null and the samples are paired with themselves, and a certain maximum extent, reached when the distance of the samples is large enough to fill the axes. At that point, even increasing the distance, the spread of the cloud will not change substantially. The spread will achieve a steady value, with small oscillation around it. The paired samples will reach their reciprocal independence.

The way the spread of the cloud varies over the distance, resumes the law of spatial continuity of the samples analysed [5]. Three analogue approaches are employed in geostatistics, to model this law:

- The correlation coefficient of the pairs, whose variation with the distance is defined as the *correlogram.*
- The covariance and the corresponding *covariance function.*
- The moment of inertia and the corresponding *variogram.*

The general definition of the moment of inertia of two paired variables x and y, follows [5]:

$$T = \frac{1}{2n} \cdot \sum_{i=1}^{n} (x_i - y_i)^2 \tag{1}$$

where the factor $\frac{1}{2}$ refers to the perpendicular distance of the n samples to the diagonal.

Hence, the empirical variogram, or semivariance, of two paired variables ($z(x)$ and $z(x+h)$ separated by the distance h is defined as follows:

$$\gamma(h) = \frac{1}{2n(h)} \cdot \sum_{i=1}^{n(h)} (z(x_i) - z(x_i + h))^2 \tag{2}$$

where the number of pairs n is represented as a function of h, because their availability changes with the distance h. The term h is usually referred as the *lag.*

A typical variogram curve reflects the empirical assumption made for the h-scatter plot. It is zero at the origin, it increases with the lag distance and starts to flatten around a certain value of variance. In Figure 2, a typical empirical variogram is shown, together with the covariance function.

For interpolation purposes, the approximation of the law of spatial continuity for all the lags is often demanded. Then, the empirical variogram is usually asked to be fitted by some theoretical analytic models.

To infer the theoretical behavior of the experimental variogram, the samples of the spatial variable are considered as the realizations of a random function, a random variable, and a series of assumptions are drawn. In particular, the assumption on the stationarity of the random function is done. In conditions of a second order stationarity [18], the empirical variogram can be conveniently fitted by a family of functions (bounded authorized models), that allow to infer the information of the spatial continuity over the entire field. Two of the most popular models used for variogram fitting are the exponential and the spherical model [5].

A direct relation between the covariance function and the variogram can be defined. The covariance function starts at the variance of the random function

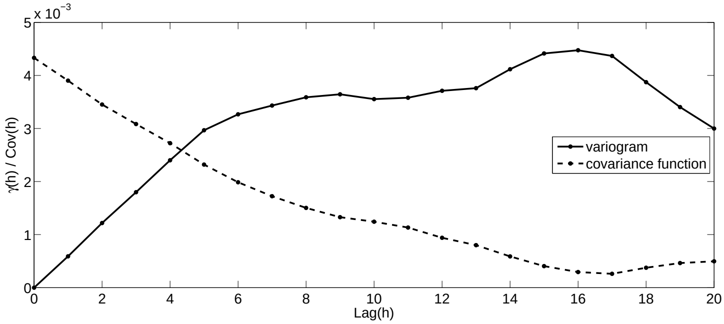


Fig. 2. A typical empirical variogram and its corresponding covariance function.

and decreases with the distance, tending to zero, when the samples of the random variable are sufficiently separated to be independent. Conversely, the variogram starts at zero, where the samples are at identical location and its variance is null, and increases with the distance, revealing the raise of the independence of the variable. It tends to the maximum degree of independence, that is the global variance of the random function [15].

The fit of the empirical variogram with the analytic models allows to parametrize the variogram function. Two main features are typically retrieved as descriptors of the shape of the theoretical variogram model: the *sill*, that is the variance at which the curve tends and the *range*, the lag value at which the sill is reached. A third very important parameter is the so called *nugget effect*. As seen, the theoretical value of the variogram at $h = 0$ is zero, because of the comparison of two different random variables at identical locations. However, in a practical experimental framework, a discontinuity of the empirical variogram at the short scale can be observed. This phenomenon is referred as the *small scale variability* [8]. The nugget effect is taken into account in the fit of the theoretical models by summing a certain quantity to the main model, such to shift the first lag to a level of variance higher than zero and cope with the small scale variability.

In Figure 3, a typical fitted theoretical model is shown, together with its main parameters.

2.2 The temporal variogram

Despite the variogram was born in a spatial statistics framework, it can be conveniently applied to time series data. Many authors [7, 4, 6] have dealt with the use of the variogram, coupled to classical signal processing techniques, as a tool for periodicity analysis of signals and time series analysis.

In the case of temporal signal processing, the distance parameter h is unidimensional and it represents the time lag among the samples. Unlike the spatial framework, where the samples are (regularly or not) distributed in the domain, in a temporal framework all the lag values are covered. The pairs availability is

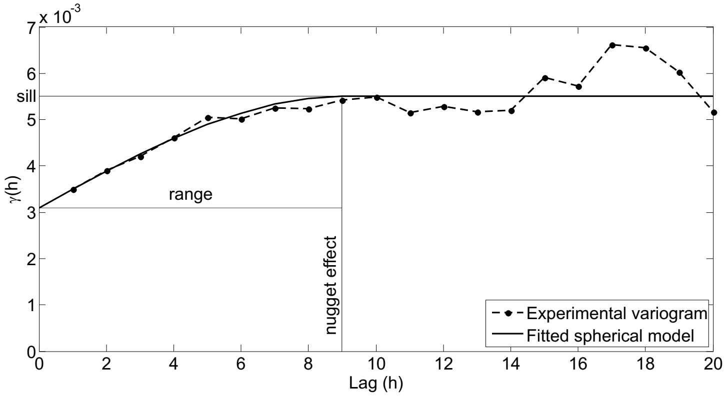


Fig. 3. An empirical variogram fitted by the analytical model, with the corresponding parameters.

a linear descending function with its maximum at lag $h = 1$, where the number of available pairs is $n - 1$ (where n represents the number of audio samples), ad its minimum, at lag $h = n - 1$, where the number of available pairs is 1.

For this reason, the reliability of the variogram values decreases with the lag. The variogram values estimated for the first lags are much more reliable than the last ones. Fortunately, the most revealing part of a variogram is indeed at the small scale, where it varies more, while a less interesting and rather constant behaviour is expressed at larger scales, just where the pairs availability decreases linearly and the estimation of this measure is less reliable.

In Figure 4, a typical temporal variogram is shown.

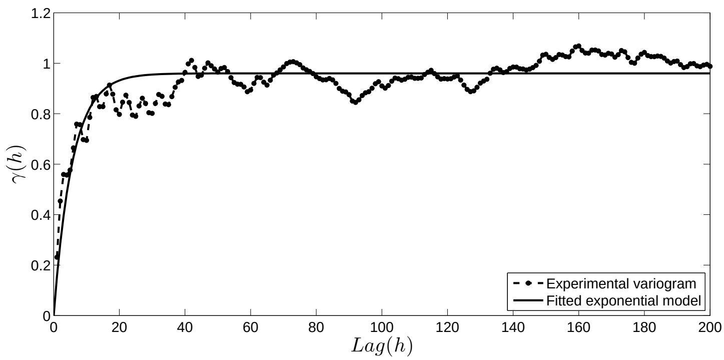


Fig. 4. A Typical temporal variogram. The experimental variogram is fitted by an exponential model.

Finally, when applied to audio signals, the variogram curve typically shows a periodical behaviour. In fact, the squared difference among the samples is affected by the periodicity of the signal itself and it is faithfully reflected by the variogram.

2.3 Variogram for MFCCs modeling

In this work, the temporal variogram is calculated on the MFCCs, as a tool for modelling the variation of the cepstral descriptor over the time fragments. A variant of the variogram proposed in [15] and a series of setups for the calculus of the distance are tested.

For the calculus of the MFCCs, the input signal is fractioned in a series of chunks with 1024 samples each, no windows overlap is employed and a hamming function is applied to each frame. The number of Mel filters (the triangular filterbank) is 40, while the number of DCT coefficients is 13. With such kind of configuration, one minute of audio signal corresponds to an MFCCs matrix of 13 x 2583 samples.

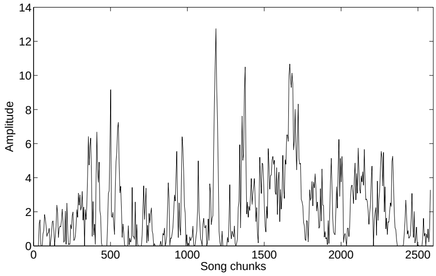
When the variogram is applied to the MFCCs, the lags values correspond to a temporal distance in terms of number of chunks in which the song has been fractioned. In order to achieve a standard measure to be employed in the quantitative comparison among the songs, each variogram is normalized by the global variance of the MFCC analysed. The result is an empirical variogram with an asymptotic tendency towards a reference variance of one. This is defined as *standardized variogram* [15].

The variogram is applied to the ISMIR 2004 Audio Description Contest (pre-MIREX) database for genre classification [2]: a set of about 700 songs, whose minimum and maximum duration was considered as 5 seconds and 5 minutes, respectively.

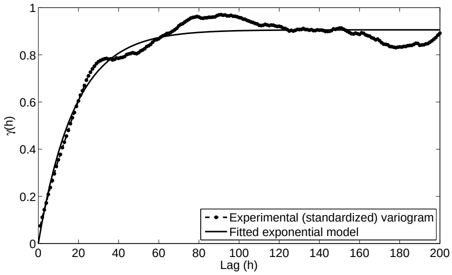
Full variogram The so called *full variogram* is the variogram of the second MFCC, calculated from lag 1 to 200. That is from the temporal pairwise distance corresponding to 1 chunk (1024 samples, about 23 ms) to the one corresponding to 200 chunks, that is about 4.6 seconds. The resulting unidimensional vector of 200 elements stands for the song signature. This approach implies a dimensionality reduction rate of about 93% (from about 2800 samples of the original MFCCs matrix (with size 215 x 13) to 200 samples of the variogram vector) in the case of the shortest audio fragment (5 seconds), and about 99.8% (from about 168000 samples of the original MFCCs matrix (with size 12919 x 13) to 200 samples of the variogram vector) in the case of the largest audio fragment with a maximum duration of 5 minutes.

In Figure 5, two examples of full variogram calculated on the second MFCC of two songs from the genre classical and electronic, are shown.

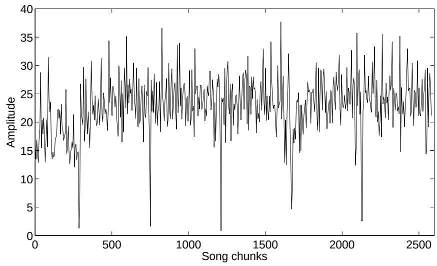
The large discrepancy expected by the comparison of two songs belonging to two very different genres, is reflected by the variogram analysis. The second MFCCs of the two songs are rather different: the one of the classical piece shows a



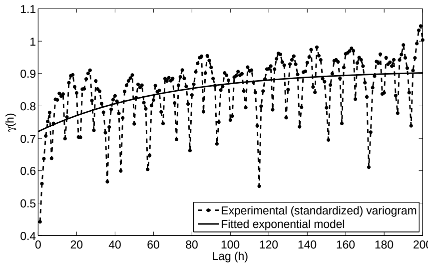
(a) The second MFCC of a classical piece.



(b) The standardized variogram of the second MFCC of a classical piece.



(c) The second MFCC of an electronic piece.



(d) The standardized variogram of the second MFCC of a electronic piece.

Fig. 5. Two examples of calculation of the full standardized variogram on the second MFCC of two songs, respectively from classical and electronic genre. The excerpts analyzed have a duration of 1 minute.

more structured and smoother variability, with few high frequency components and a hidden (or missing) periodicity, while the one of the electronic piece is much more fuzzy, with a large contribution of a high frequency variability and a marked periodical behaviour.

The corresponding variograms reflects very well the behaviour highlighted. The variogram of the classical piece reveals a very structured variability, with a high pairwise continuity at the small scale (the nugget effect is null) and a smoothly increasing variance with a clear asymptotic trend towards the range. Conversely, the variogram of the electronic piece is much more unstructured, with continuous periodic oscillation coupled to a very weak asymptotic trend. Its nugget effect is rather high.

Reduced variogram The *reduced variogram* is calculated on 12 MFCCs (from the second MFCC to the last one), on a reduced bunch of lags. A total amount of 20 lags are sampled with a logarithmically varying density, from 1 to 200, with the aim to concentrate the lags at the smallest scale, where most of the variance is expressed (see Figure 6).

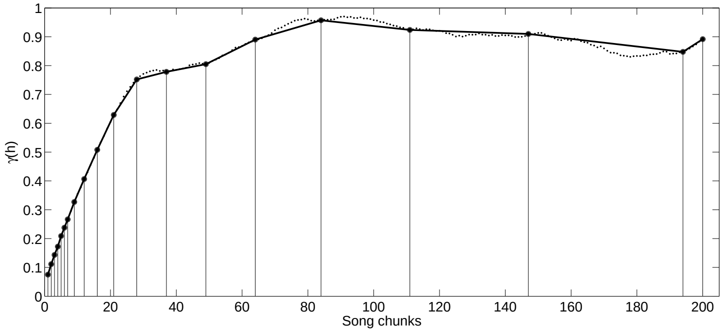


Fig. 6. The variogram for the classical piece of Figure 5(b), reduced by the lag sampling (thick line). Note the logarithmic distribution of the sampled lags.

The signature matrix is of size 12 x 20, resulting in a total amount of 240 elements (if stacked). The dimensionality reduction rate is quite the same of the full variogram. In Figure 7, the reduced versions of the full variogram of Figure 5 are shown.

The conclusions drawn for the reduced matrix of variograms are the same as for the full variogram. The classical piece shows smoother variograms, revealing a more structured variability and a high small scale pairwise continuity. Conversely, the electronic track reveals a more fuzzy variability structure and a marked periodical behaviour. In both cases, the reduction of the number of the lags keeps guaranteeing a faithful representation of the original full variogram.

2.4 Distance measurement

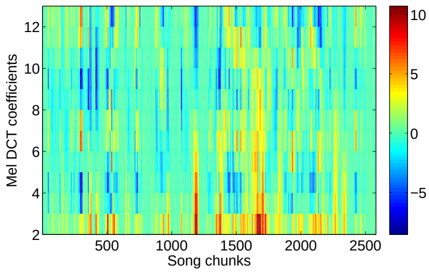
In order to estimate the degree of similarity of the songs, the signatures have to be numerically compared. In this work, a weighted Euclidean distance is used.

In general, the distance is calculated as follows:

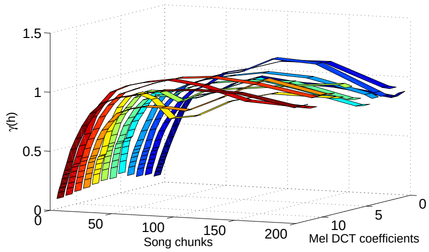
$$D_{i,j} = \sqrt{\sum_{k=1}^{n} \left((V_i(k) - V_j(k)) \cdot \omega(k) \right)^2} \tag{3}$$

where $V_i(k)$ and $V_j(k)$ are the values of the k-th lag of the variograms of two songs i and j, and $\omega(k)$ is the weight of the k-th lag, with a maximum number of lags n equal to 240, for a bi-dimensional reduced variogram and 200, for a full unidimensional variogram. Note that the bi-dimensional variogram is stacked into a unidimensional vector to simplify the calculus.

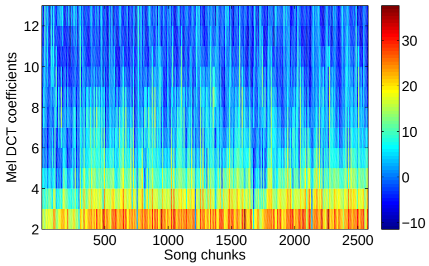
Actually, the variogram shows a maximum of information (in term of quality and reliableness) at the small scale. The most predominant meaning of the measure arises from the first lags, up to the achievement of the range, beyond which the variogram looses significance. For this reason, three different sets of weights



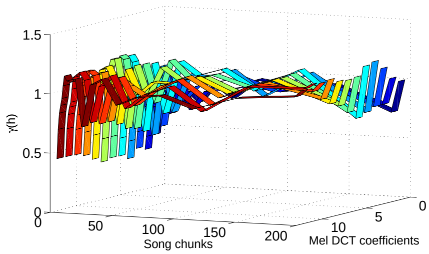
(a) The whole MFCCs matrix of a classical piece.



(b) The matrix of standardized variograms of the whole MFCCs matrix, reduced by the lags sampling. Classical piece.



(c) The whole MFCCs matrix of a electronic piece.



(d) The matrix of standardized variograms of the whole MFCCs matrix, reduced by the lags sampling. Electronic piece.

Fig. 7. Two examples of calculation of the matrix of standardized variograms of the whole MFCCs matrix of two songs, respectively from classical and electronic genre. The excerpts analyzed have a duration of 1 minute.

are proposed: a set of exponentially decreasing weights, a set of logarithmically decreasing weights and, finally, a set of linearly decreasing weights. A fourth unweighted variant of the distance is included.

In Figure 8, the three sets of weights are compared. Note that the vectors of weights represented here correspond to one of the stacked vectors of weights employed for the reduced variogram (20 lags).

Note that in any of the three cases, the weights are normalized such that their sum is 1.

3 Evaluation of the performance of the algorithms

The evaluation of the performance of the two variants of variogram, each with the four weighting functions, is implemented on the basis of the genre classification music database of the ISMIR 2004 Audio Description Contest [2].

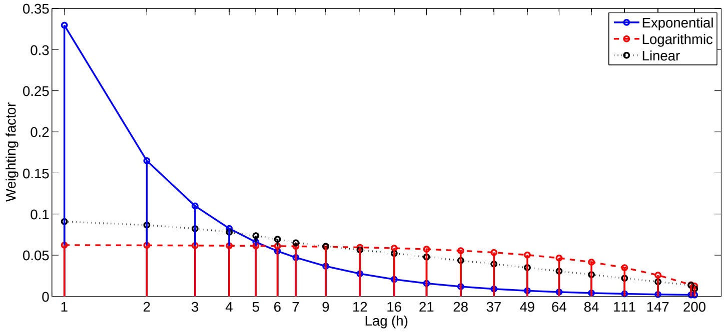


Fig. 8. The three vectors of weights employed for the calculus of the distance. Note that the shape of the linear weights is deformed by the scaling of the lag axis that is logarithmic.

The pseudo-objective evaluation [3], currently employed in the MIREX music similarity tasks, is performed. The matching rates of artist, album and (artist-filtered) genre, for the first 5, 10, 20 and 50 songs are calculated.

After sorting the list of songs according to the degree of similarity to the seed item (one of the songs of the collection, selected recursively), the pseudo-objective statistics are calculated as percentages of the songs of the list sharing the same artist, album or (artist-filtered) genre. These percentages are calculated four times, on a reference total of the first 5, 10, 20 and 50 songs of the list.

In order to compensate for the unequal distribution of items per category (artist, album or genre), the reference total is defined as the maximum between the defined reference (5, 10, 20 or 50) and the maximum number of available songs per category. For instance, if only 8 songs are available for a certain artist, the reference total for the calculus of the artist-based statistic has to be 5, for the first 5 songs, but it must be reduced to 8 for each of the higher counts (10, 20 or 50). In fact, the statistics would be negatively affected by considering the reference total as some values higher than the maximum allowed by the database itself. If the algorithm is able to return all the 8 correct correspondences for the artist into the first 8 positions, it has to be considered as best performing for each of the totals: 5/5 (for the first 5 items) and 8/8 (for the first 10, 20 or 50 items).

This procedure is recursively applied to all the songs of the collection, setting each time one of them as the seed song. Finally, the global score is calculated as the averaged mean of the scores obtained for each seed song. In Table 1, the matching scores for the two variants of variogram are shown.

The performance returned by the reduced variogram is globally better than the one of the full variogram, for any kind of weighting configuration. On the one hand, it is true that the full variogram returns a more complete information of the second (and most representative) MFCC, with respect to the reduced

	Full variogram				Reduced variogram			
	Exponential weights							
Artist	7.24	9.12	13.01	23.63	16.99	18.47	23.11	34.26
Album	5.29	8.52	14.23	26.20	13.25	19.14	26.30	37.69
Genre	**43.62**	**42.46**	**41.85**	**40.16**	**46.43**	**45.29**	**44.17**	**43.23**
	Logarithmic weights							
Artist	5.20	5.94	9.37	18.05	15.81	16.70	21.88	32.86
Album	3.86	5.70	10.43	19.40	12.62	17.83	25.72	37.03
Genre	**38.72**	**38.48**	**38.41**	**37.65**	**48.07**	**46.66**	**45.51**	**42.84**
	Linear weights							
Artist	5.14	5.81	8.90	18.45	16.51	17.70	22.40	33.77
Album	3.79	5.47	9.64	19.44	12.93	19.15	26.74	38.68
Genre	**39.25**	**38.44**	**38.40**	**37.79**	**47.84**	**46.94**	**44.81**	**42.73**
	No weights							
Artist	4.50	5.47	8.40	16.43	15.23	16.32	21.59	32.38
Album	3.42	5.65	9.23	17.97	13.03	17.55	25.23	36.07
Genre	**38.19**	**38.08**	**37.63**	**37.25**	**48.48**	**47.68**	**46.18**	**43.70**
	First 5	**First 10**	**First 20**	**First 50**	**First 5**	**First 10**	**First 20**	**First 50**

Table 1. Pseudo-objective statistics for the two variants of variogram calculation. Note that the genre scores are calculated on the artist-filtered subset. The genre results are in bold because these results are the ones that can be compared with the MIREX AMS 2011 results presented in table 2. It can be observed that the proposed methods perform quite well.

variant, that is calculated on a smaller bunch of lags. On the other hand, the completeness of the information based on the involvement of the complete set of MFCCs returns a more accurate description of the song analysed. Apparently, the loss of information due to the reduction of the lags is compensated by the gain derived by the employment of the complete MFCCs matrix.

Also, an inverse trend of variation of the scores is observed for the three different categories: artist, album and genre. In particular, the artist and album-based scores increase with the number of items considered, while for the genre the tendency is inverse. It basically depends on the availability of items per category. In fact, the probability of returning one song of the first 5, 10, 20 or 50 with the same genre of the seed song, is much higher than the one related to the other two categories. Actually, the genre-based statistic reflects the higher concentration power of the genre, that finds similar songs more easily, yet from the very first few items considered. Conversely, as finding the correct songs with the same artist or album is much harder, the larger the number of items considered, the higher the score obtained for these categories.

The best results for the full variogram (e.g.: 43.62%, obtained for the genre coincidence of the first 5 items of the list) have been obtained with the set of exponentially decreasing weights, where the first lags contribution is much higher than the others. Surprisingly, the result shown by the reduced variogram is different: the best scores are referred to the null weighting of the distance, although the trend is not as clear as the case of the full variogram.

Table 2 shows the results of the pseudo-objective evaluation of the algorithms proposed to the Audio Music Similarity contest of the MIREX 2011 (only the artist-filtered genre scores) [12]. It can be observed that the scores obtained for the variogram-based approaches are in line with the reference represented by the MIREX Audio Music Similarity task. Although a direct quantitative comparison cannot be provided because of the differences in the test database used in the two frameworks, the variogram seems to return a rather reliable accuracy in the estimation of music similarity.

Method	First 5	First 10	First 20	First 50
STBD1	24.19	23.34	22.14	20.57
STBD2	23.55	22.56	21.61	19.98
STBD3	23.07	22.55	21.78	20.47
DM2	46.02	44.14	42.22	39.28
DM3	46.08	44.20	42.33	39.37
GKC1	23.45	22.55	21.57	20.01
HKHLL1	34.91	33.81	32.72	31.39
ML1	41.77	39.86	38.09	35.53
ML2	40.19	38.45	36.28	33.62
ML3	41.06	38.99	36.80	33.85
PS1	54.11	52.17	50.13	46.74
SSKS3	54.65	53.15	51.52	48.98
SSPK2	54.24	52.75	51.19	48.56
YL1	37.40	35.43	33.01	29.54

Table 2. Average artist-filtered genre scores of the algorithms proposed to the MIREX 2011 contest. The method acronyms correspond to the standard coding employed in the MIREX contest [12].

4 Conclusions and future works

In this paper, the use of the temporal variogram has been proposed as a tool to model the temporal variability of the Mel Frequency Cepstral Coefficients and it has been exploited to estimate music similarity.

After a brief description of the theory of the variogram analysis and its adaptation to a temporal framework, two different variants of the calculus of the variogram and four weighting functions for the calculus of the distance between the song signatures, have been proposed. Both the two variogram-based approaches have been tested on a reference database of songs, divided into six different genres. A pseudo-objective analysis has been computed in order to achieve a quantitative evaluation of the performance of the methods and propose a discussion on the results. Also, a comparison with the actual reference in term of algorithms aimed to perform music similarity, has been provided.

The reduced variogram returns better scores than the full variant, due to the more complete information given by the whole MFCCs matrix (the first MFCC excluded). This method seems not be really influenced by the kind of weighting function used for the calculus of the distance. The results are in line with the references of the MIREX 2011.

All the variograms analyzed so far are the results of the empirical calculus of the equation (2). In future development of the variogram-based approach, the automatic fitting of the theoretical models can be employed to try to resume the variogram function as a series of parameters. In particular, the nugget effect, the range and the sill of the theoretical models could be employed as low-level descriptors for classification purposes.

In order to test this concept, a very simple approximation has been carried out. A simple least square fit of the exponential model has been implemented to the variograms of the songs of the collection tested in the article, in order to achieve a first estimation of the nugget effect. Afterwards, it has been employed as a low-level descriptor, together with other popular MIR descriptors [16], and tested in a music genre classifier. The classifier employed was a simply knn-classifier, with $k = 5$ neighbors.

The results are rather encouraging. The performance of the nugget effect, although it has been estimated by a simply automatic fit of the experimental variograms, are in line with the one of other more popular features. The nugget effect reflects even a better behavior in some specific cases (as the example of the genre world).

As known, the automatic fitting of the empirical variograms is an actual matter of discussion and the issue is far from being resolved [9]. These preliminary results presented encourage to focus on the automatic fitting of the variogram models in order to obtain more robust descriptors to be conveniently used for MIR tasks.

Acknowledgments. This work has been funded by the Ministerio de Economía y Competitividad of the Spanish Government under Project TIN2010-21089-C03-02 and by the Ministerio de Industria, Turismo y Comercio of the Spanish Government under Project TSI-090100-2011-25.

References

1. Aucouturier, J.J., F., P.: Improving timbre similarity: How high is the sky? Journal of Negative Results in Speech and Audio Sciences 1(1) (2004)
2. Cano, P., Gmez, E., Gouyon, F., Herrera, P., Koppenberger, M., Ong, B., Serra, X., Streich, S., Wack, N.: Ismir 2004 audio description contest (2006), `http://mtg.upf.edu//files/publications/MTG-TR-2006-02.pdf`
3. Downie, S.J.: The music information retrieval evaluation exchange (mirex). D-Lib Magazine 12(12) (2006)

4. Haslett, J.: On the sample variogram and the sample autocovariance for non-stationary time series. The Statistician 46(4), 475–485 (1997)
5. Isaaks, E.H., Srivastava, M.R.: An Introduction to Applied Geostatistics. Oxford University Press, USA (January 1990)
6. Kacha, A., Grenez, F., Schoentgen, J., Benmahammed, K.: Dysphonic speech analysis using generalized variogram. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. (ICASSP '05). vol. 1, pp. 917–920 (2005)
7. Khachatryan, D., Bisgaard, S.: Some results on the variogram in time series analysis. Quality and Reliability Engineering International (March 2009)
8. Krige, D.G.: A statistical approach to some basic mine valuation problems on the witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa 52(6), 119–139 (December 1951)
9. Li, S., Lu, W.: Automatic fit of the variogram. In: Third International Conference on Information and Computing (ICIC), 2010. vol. 4, pp. 129 –132 (june 2010)
10. Logan, B., Salomon, A.: A music similarity function based on signal analysis. In: Multimedia and Expo, 2001. ICME 2001. IEEE International Conference on. pp. 745 – 748 (aug 2001)
11. Mandel, M.I., Ellis, D.P.W.: Song-Level Features and Support Vector Machines for Music Classification. In: Reiss, J.D., Wiggins, G.A. (eds.) Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR). pp. 594–599 (Sep 2005)
12. MIREX2011: Audio music similarity and retrieval results (2011), `http://www.music-ir.org/mirex/wiki/2011:Audio_Music_Similarity_and_Retrieval_Results`
13. Pampalk, E.: Computational Models of Music Similarity and their Application to Music Information Retrieval. Ph.D. thesis, Vienna University of Technology, Vienna (March 2006)
14. Rabiner, L., Juang, B.H.: Fundamentals of speech recognition. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1993)
15. Sammartino, S., Tardón, L.J., de la Bandera, C., Barbancho, I., Barbancho, A.M.: The standardized variogram as a novel tool for music similarity evaluation. In: Proc. of Int. Symposium on Music Information Retrieval (ISMIR 2010). pp. 559–564 (2010)
16. Tardón, L.J., Sammartino, S., Barbancho, I.: Design of an efficient music-speech discriminator. Journal of the Acoustical Society of America 127(1) (2010)
17. Tzanetakis, G., Cook, P.: Musical genre classification of audio signals. Speech and Audio Processing, IEEE Transactions on 10(5), 293 – 302 (jul 2002)
18. Wackernagel, H.: Multivariate Geostatistics: An Introduction With Applications. Springer-Verlag Telos (January 1999)