THE STANDARDIZED VARIOGRAM AS A NOVEL TOOL FOR AUDIO SIMILARITY MEASURE

Simone Sammartino, Lorenzo J. Tardón, Cristina de la Bandera, Isabel Barbancho, Ana M. Barbancho
Dept. Ingeniería de Comunicaciones, E.T.S. Ingeniería de Telecomunicación,
Universidad de Málaga, Campus Universitario de Teatinos s/n, 29071, Málaga, Spain
ssammartino@ic.uma.es

ABSTRACT

Most of methods for audio similarity evaluation are based on the Mel frequency cepstral coefficients, employed as main tool for the characterization of audio contents. Such approach needs some way of data compression aimed to optimize the information retrieval task and to reduce the computational costs derived from the usage of cluster analysis tools and probabilistic models. A novel approach is presented in this paper, based on the standardized variogram. This tool, inherited from Geostatistics, is applied to MFCCs matrices to reduce their size and compute compact representations of the audio contents (song signatures), aimed to evaluate audio similarity. The performance of the proposed approach is analyzed in comparison with other alternative methods and on the base of human responses.

1. INTRODUCTION

The computation of the degree of similarity among songs is one of the most demanded tasks in the field of multimedia processing, and its interest is still growing with the increasing popularity of on line services and databases. Music Information Retrieval (MIR) stands for the tools to access audio contents with the aim to reorder, search and classify them [1]. In MIREX 2006 [3], the term 'Audio Similarity' was introduced for the first time in the tasks list and, consequently, a human evaluation system (Evalutron6000) was created to make quantitative evaluations of the proposed algorithms.

The main task of audio similarity evaluation is based on the definition of some form of representation of the songs (signatures) to compare them and measure the closeness of the signature songs. The base of most of the known algorithms for audio similarity evaluation are the Mel frequency cepstral coefficients (MFCCs) [10]. The spectral information supplied by the MFCCs is proposed to be compressed in a wide variety of different approaches by different authors [13] [4] [11] [1]. In this work, the standardized variogram [8] is presented as a novel tool to conveniently compress the MFCCs vectors for sorting similar songs.

The outline of the paper follows: in Section 2, a general description of the MFCCs is presented. In Sections 3 and 3.2, the details of the approach based on the variogram and its application to signal processing are described. In Section 4, the use of MFCCs matrices is presented and in Section 5 the application of the variogram is discussed in detail. Finally, in Section 6, the experimental results are presented and in Section 7, the conclusions and future proposals are discussed.

2. MEL COEFFICIENTS AND AUDIO SIMILARITY

The MFCCs are short-term spectral-based features, originally developed for speech recognition and successfully adapted to music information retrieval [10]. The computation of MFCCs follows some crucial steps [13]: 1) the calculus of the short-term spectrum of the signal, 2) the transformation of the spectrum into the Mel scale (through a triangular filter bank), 3) the calculus of the logarithm of the Mel spectrum and 4) the compression of the resulting matrix through the application of the DCT (Discrete Cosine Transform). MFCCs are widely used to generate compact spectral representations of the song: the signal is framed into short fragments (usually some tens of milliseconds) and their coefficients are computed frame by frame [13]. In order to conveniently represent the global spectral behavior of the song in a compact way, the MFCCs vectors have to be clustered. For this task, several approaches have been proposed by different authors. Pampalk [13] uses GMM and EM approach, by modelling the probability distribution functions of the coefficients vectors. Foote [4] proposes a supervised tree-structured quantizer as discriminant approach for the sequential labeling of the coefficients. Aucouturier and Pachet [1] present a combination of GMM/EM and Monte Carlo approaches to evaluate the likelihood between the MFCCs of two different songs. Finally, Logan and Salomon [11] propose the popular K-means method for MFCCs clustering. In this article, an alternative approach is proposed based on the computation of the variogram of the MFCCs, allowing for a computationally low-cost compression of the coefficients and a simple calculus of the distance among the spectral signatures.

3. THE STANDARDIZED VARIOGRAM

The term 'variogram', inherited from Geostatistics, stands for the function describing the evolution of the spatial dependence of a random field [16]. Empirically used by the mine engineer D.G. Krige in South Africa mines [9], and later formalized by G. Matheron in its pioneer works [12], the variogram or 'semivariance function' is widely employed in spatial statistics to perform uncertainty modeling in a spatial framework. It is often used as characteristic weighting function for the spatial interpolation technique known as *Kriging* [9].

3.1 Some mathematical issues

A formal definition of the variogram is now provided. Let z_α, with $\alpha = 1, \cdots, n$ represents a set of n sampled observations of a spatial phenomenon. The variogram is defined as half the variance of the increment $[z_\alpha - z_{\alpha+h}]$ [16]:

$$\gamma(\alpha, h) = \frac{1}{2}E\{[z_\alpha - z_{\alpha+h}]^2\} - \{E[z_\alpha - z_{\alpha+h}]\}^2 \quad (1)$$

Assuming the intrinsic stationarity of order two [16], the mean of the variable $E[z]$ is invariant for any translation, that is $E[z_\alpha] = E[z_{\alpha+h}]$, the second term of equation (1) can be neglected and the variance of the increment is said to be depending only on the distance vector h and not on the position α [8]:

$$\gamma(h) = \frac{1}{2}E\{[z_\alpha - z_{\alpha+h}]^2\} \quad (2)$$

where z_α and $z_{\alpha+h}$ are two different samples of the random variable z separated by a distance h.

Given a set of spatially distributed samples, the variogram can be estimated empirically [16]:

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N_h} [z_\alpha - z_{\alpha+h}]^2 \quad (3)$$

where the number of pairs $N(h)$ depends on the value of h. For its mathematical relation with the variance, the variogram is also known as semivariance function (or semivariogram).

The variogram is strictly related with the auto-covariance function of the increment. In particular the covariance of the increment $Cov(z_\alpha, z_{\alpha+h})$, in condition of translation invariance of the mean, can be expressed as follows:

$$Cov(z_\alpha, z_{\alpha+h}) = Cov(h) = E[z_\alpha \cdot z_{\alpha+h}] - E[z_\alpha]^2 \quad (4)$$

where $E[z_\alpha] = E[z_{\alpha+h}]$. When h is zero, $Cov(0)$ is maximum and it corresponds to the variance of the variable:

$$Cov(0) = E\{[z_\alpha]^2\} - \{E[z_\alpha]\}^2 = Var(z_\alpha) \quad (5)$$

Note that equation (2) can be written as:

$$\gamma(h) = \frac{1}{2}E\{[z_\alpha]^2 - 2 \cdot z_\alpha \cdot z_{\alpha+h} + [z_{\alpha+h}]^2\} \quad (6)$$

and using equations (4) and (5), we can express the variogram in term of the covariance function:

$$\gamma(h) = Cov(0) - Cov(h) \quad (7)$$

The last equation shows the relation between the variogram and the covariance function [14]. Under the condition of translation invariance of the mean, at $h = 0$, the covariance is just the variance of the variable, $Cov(0) = Var(z)$, and the variogram is zero, $\gamma(0) = 0$. Conversely, when the pair of elements, z_α and $z_{\alpha+h}$, are too far away to show any kind of relation, their covariance is zero and the variogram is the variance of the variable, $\gamma(h) = Var(z)$. In general, the covariance function shows a behavior opposed to the behavior of the variogram (see Fig. 1).

The empirical variogram is usually fitted by a theoretical model to obtain a continuous function, modeling the covariance exhaustively in the whole domain. The models are chosen within a group of *admissible models* that must be positive-definite [6]. Moreover, the theoretical models can be characterized by few shape parameters [6]: the *Sill*, the asymptotic variance value the function tends to when the lag distance, h, increases, the *Range*, the lag value at which the theoretical variogram reaches the sill, and the *Nugget effect*, the discontinuity of the function at the origin.

When semivariance values are normalized by the global variance, the variogram is reported as *standardized variogram* [8] and its correspondence covariance function is the correlation function. Both the empirical and its correspondent theoretical standardized variogram are shown in Fig. 1. The correspondent correlation function is shown too.

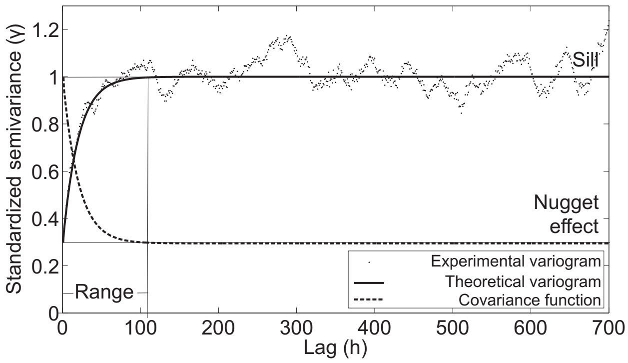


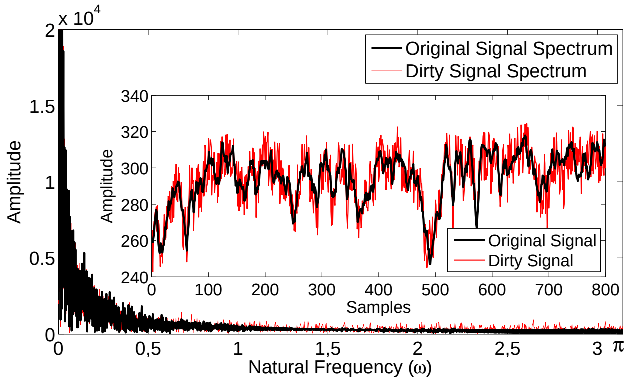
Figure 1. An example of a typical standardized variogram. The empirical variogram (dotted line) is fitted by the theoretical model (solid line). The correlation function (dashed line) is shown too.

3.2 The variogram in signal processing

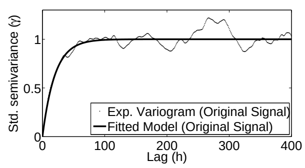
Many authors have dealt with the use of the variogram coupled to classical signal processing techniques, as a tool for periodicity analysis of signals and time series analysis. Khachatryan and Bisgaard [8] employ the variogram as tool for estimating the stationarity of industrial time series data. Haslett [5] proposes the use of the variogram as a functional approach for time estimation in case of fault of the stationarity conditions. Kacha et al. [7] apply the generalized variogram to the linear prediction in disordered speech analysis.

In spite of the different origins of the spatial variographic approach and the time series analysis in signal processing, the former can be successfully applied as an alternative tool for spectral analysis. In the case of time-signal processing, the parameter h is unidimensional and it represent the time lag among the samples.

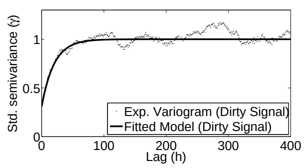
If we take a signal and we add an uncorrelated noise component with known mean and variance (see box inside Fig. 2(a)), we can observe that it is well reflected both in the waveform and in the frequency spectrum. In the variogram, the added signal leads to a very small change in the general shape of the curve (Fig. 2(b) and 2(c)), while a marked increase of the variance at the origin (nugget effect) is noticeable. Such value corresponds to a contribution of about the 31% of the total variance of the dirty signal. However, in the frequency spectrum, this change is mainly reflected in central-high frequency bands, where only a diffused increase in amplitude is apparent (Fig. 2(a)).



(a) Spectra of original (darker thick line) and dirty signal (lighter thin line). Inner box: raw signal.



(b) Standardized Variogram of original signals.



(c) Standardized Variogram of dirty signals.

Figure 2. Spectra and standardized variograms of a clean signal and of the same signal corrupted with additive noise. Experimental variograms are represented with a dotted line and the theoretical fitted model with a solid line.

The variogram can also be conveniently used as tool for sound analysis applications. Dillon et al. [2] noted as the variogram fluctuations, once reached the sill, are strictly related with signal spectrum. He remarks that the variogram can be especially useful for fundamental frequency detection, by taking into account the variance pseudoperiodic pattern known as *hole effect* [6].

4. COMPRESSION OF MATRICES

The standardized variogram described in Section 3 is proposed here as a novel method for reducing the dimensionality of the MFCCs matrices. It is computed on each vector of coefficients throughout the frames of the song, to obtain a function describing the evolution of the covariance through the time. With the aim to compress the MFCCs information, the variogram is computed only on a reduced number of lags (values of distance h for which the variogram is calculated). As shown in Fig. 1, the variogram typically presents a logarithmic-like rising behavior at the lowest lags and an asymptotic trend to the global variance (equal to one in the case of standardized variogram) from lags approaching the range, forward. Taking into account these two factors, a total amount of ten lags values are sampled logarithmically from 1 to half the length of Mel coefficients.

For each row of the MFCCs matrices, the semivariance is computed for all the pairs of samples located at distances equal to the lags selected, and the values are normalized by the variance of the MFCCs row data. The outcome is a compact function keeping enough information to characterize the signal.

The experimental standardized variogram can be characterized on the basis of two parameters extracted from its correspondent theoretical function (although the latter is not explicitly calculated in this application): the range and the nugget effect. In this case, these parameters can be interpreted on the base of their spectral meaning.

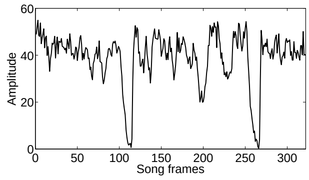
The range can be interpreted as the time scale at witch the periodicity of the signal begins to be evident. Up to the range, the structured variability of the variable masks its periodicity, while, when the pairwise covariance starts to be weak enough (from range forward), that periodic behavior rises and it becomes evident. Clearly, due to the strong reduction of lags, sometimes the range can be poorly detected by the reduced variogram.

The nugget effect is very important to understand the small-scale behavior of the Mel coefficients. The discontinuity at lag $h \to 0$ explains the variation of the signal at very small-scale. In terms of spectral analysis, it stands for the high frequency contribution to the total variance in the Mel spectra.

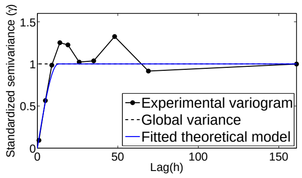
An example of the application of the variogram to the MFCC spectra is shown in Fig. 3. The first song (Fig. 3(a)) is a piece from the genre 'Classic' [3], its spectrum shows a rather clear periodicity, with some peculiar patterns repeating with a certain regularity. The second song, belonging to the genre 'Heavy metal', shows a more fuzzy spectrum with higher frequency variations and a less evident periodicity. Such differences are well reflected in their correspondent variograms.

The clearer regularity of the classic piece is reflected by a certain degree of periodicity in the variogram (although not exhaustively revealed by the reduced number of lags). Moreover, the very low nugget variance value reflects the high degree of regularity with reduced high frequency oscillations (Fig. 3(b)).

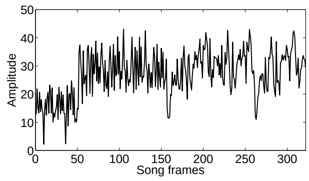
In the case of the heavy metal piece, the very high frequency oscillations (Fig. 3(c)) are well reflected by a notable nugget variance (about the 50% of the total variance) and by a larger range indicating the lack of a structured



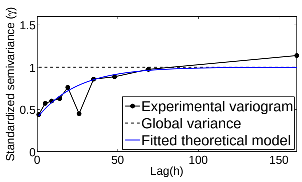
(a) The second Mel coefficient of a classical music song.



(b) Standardized reduced variogram of the signal in fig. 3(a) (dotted line).



(c) The second Mel coefficient of a heavy metal song.



(d) Standardized reduced variogram of the signal in fig. 3(c) (dotted line).

Figure 3. A comparison between the variograms of the MFCCs spectra of two very different songs, a classical music song (top) and a heavy metal song (bottom). Fitted theoretical model (thick lighter line) and global variance (dashed line) are shown too. The pieces are 35 seconds length.

variability. Note that, although poorly reflected by the reduced availability of lags, some degree of non-stationarity, expressed as the lack of a well definite asymptoticity of the variogram [16], is present in this case.

5. STANDARDIZED VARIOGRAM FOR AUDIO SIMILARITY ASSESSMENT

As mentioned before, the variogram has been employed for audio similarity assessment. For each piece, the Mel coefficients are calculated and the standardized variogram is computed for each coefficient, obtaining a compact signature of the track. Successively, the signatures are compared, by computing a weighted difference of their elements.

After computing the standardized variogram (10 lags) of the 19 Mel coefficients for each song (the first one has been neglected [13]), the resulting 10×19 matrices (signatures) are compared by averaging the weighted absolute value of their difference, according to the following equation:

$$D = \frac{1}{I \cdot J} \sum_{j=1}^{10} \sum_{i=1}^{19} \omega_j \left| V_a(i,j) - V_b(i,j) \right| \qquad (8)$$

where V_a and V_b are the signature matrices for song a and b, respectively and the indexes i and j are referred to the 19 coefficients and the 10 lags, respectively. Differences are linearly weighted in order to give more importance to the small-scale lags of the variogram vectors. The vector $\Omega = [\omega_1, \ldots, \omega_{10}]$ contains the 10 linearly decreasing weights ω_j, such that $\sum_{j=1}^{10} \omega_j = 1$. Each j-th weight is computed as follows:

$$\omega_j = \frac{11 - j}{D} \qquad (9)$$

where $D = \sum_{j=1}^{10} j$. Audio similarity is simply evaluated by sorting the songs with respect to a reference piece, according to their reciprocal distance, computed using equation (8).

6. EXPERIMENTAL RESULTS AND DISCUSSION

An objective evaluation of the sorting capability of the method is very hard to achieve because of the subjectiveness of the concept of 'audio similarity'. Actually, one song can be judged as more similar to another one depending on a series of parameters (rhythm, spectral content, melody etc.) that are subconsciously evaluated by the listeners.

In order to obtain a robust and objective estimate of the performance of the method, a series of tests performed by a group of users, have been carried on. A total of 5 lists of songs have been submitted to 10 users who sorted them with respect to a set of reference songs, without any previous knowledge about any tagging or taxonomy of the dataset. The test songs are sampled by the Audio Description Context database of the ISMIR2004 [3] and belong to all the genres presented in the database.

Successively, the lists created manually have been compared with the outcomes of 4 automatic methods, the variogram-based method and other three methods that can be found in the literature:

1. Fluctuation Patterns [13]
2. MFCCs with GMM/EM clustering approach [13]
3. MFCCs with K-means clustering approach

The fluctuation patterns, describing the amplitude modulation of the loudness of the frequency bands, are used by Pampalk [13] to briefly characterize the song spectral content. The Gaussian Mixture Models coupled with Expectation/Maximization approach are employed by the same author to cluster the Mel coefficients in 30 vectors (G30) of 19 elements. The third method is the same approach used by Logan and Salomon [11], based on the calculus of the MFCCs clustered by the popular K-means, with the Euclidean distance instead of the Kullback-Leibler distance.

A total amount of some tens of lists have been obtained by the manual sorting by the users. A rapid look at these lists reveals a strong lack of homogeneity among them. It is related to the high subjectiveness of the sorting process and the variability of the human perception of the 'audio similarity'. This leads to the lack of a representative list for each reference song. Instead of trying to extract a unique reference list among the users, the authors turned to derive a measure of the agreement among the users.

A weighted matching score has been computed, taking into account the reciprocal distance of the songs (in terms of position index in the list). Such distances have been linearly weighted, such that the first songs in the lists reflected more importance than the last ones. Actually, it is easier to define the order of few very similar songs, than to sort the very different ones.

Let L_α and L_β represent two different lists of n songs, for the same reference song, the matching score S has been computed using the following equation:

$$S = \sum_{i=1}^{n} |i - j| \cdot \omega_i \quad (10)$$

where i and j are the indexes for lists L_α and L_β, respectively. In particular, j is the index of the j-th song in list L_β, such that $L_\alpha(i) \equiv L_\beta(j)$. In practice, the i-th song in the list L_α is searched in L_β and their correspondent indexes are compared. The absolute difference is linearly weighted by the weights ω_i as referred in equation (9).

Finally, the scores are transformed to be represented as percentage of the maximum score attainable.

For each reference song, the matching scores have been computed among all the available lists, both among the users lists and among the users lists and the ones returned by the automatic methods. Thus, two different sets of scores have been obtained: the inter-users scores and the users-automatic scores. The measure of the performance of the automatic method is drawn by the degree of similarity of the two sets, that is, how close are the scores computed among the users lists and the lists returned by the automatic method. In order to have an estimation of such closeness, the coherence among the two sets of scores is computed by a statistical test. The Kolmogorov-Smirnov test [15] has been used to measure the correspondence between the two distributions of the two sets of scores, before and after the inclusion of the automatic list.

In Table 1, the basic statistics for both the distributions of the inter-users scores set and the users-automatic scores sets are shown. The results of the statistical test (H) is shown too.

The degree of similarity among the songs is a very subjective response and only a high number of cases can guarantee a reliable response. Nevertheless, the statistical results are enough to have an idea of the performance of the automatic methods. The response of the users can be seen as some form of quantifying the difficulty level of the sorting task. When the songs are easily sortable, the users show a high degree of agreement (high mean scores). Conversely, when the similarity among the songs is not very clear, the discrepancy among the users increases and, together with a decrease of the centrality measures (mean and median), an increase of the variance is appreciable. Actually, the standard deviation is an index of the disagreement among the users and can be related with the complexity of the sorting procedure.

In the test results, the discordance among the users is well reflected by high values of the standard deviation in most of the cases. The mean standard deviation for the 5 cases is about the 14% of the mean score.

In general, a wide variety of performances are shown by the different methods. The method based on the clustering of the Mel coefficients by the GMM/EM approach reaches the best score in 3 cases, for songs B,C and D, while it fails

Ref.song	Method	Mean	Median	Min	Max	Skewness	St.Dev.	H
Song A	**Users**	72.3	74.9	33.4	90.9	-1.0	13.2	-
	MFCC-Var	**71.5**	**75.2**	42.8	82.5	-1.8	11.3	0
	FP	71.2	72.5	**43.5**	**85.3**	-1.3	11.6	0
	MFCC-EUC	70.8	71.8	41.8	84.8	-1.4	12.1	0
	MFCC-G30	55	54.3	41.3	69.4	0.1	**9**	1
Song B	**Users**	81.8	83.9	52.7	99.2	-1	9.6	-
	MFCC-Var	75.4	76.8	54.4	87.8	-1.2	8.7	1
	FP	66.6	66.1	58	81.8	0.9	6.9	1
	MFCC-EUC	67.5	66.6	**61.5**	72.9	0.1	**4.3**	1
	MFCC-G30	**76.7**	**82.3**	41.5	**99.2**	-0.7	14.2	1
Song C	**Users**	84.3	85.6	66.6	96.2	-0.2	6.6	-
	MFCC-Var	71.3	70.8	66.1	76.2	0.2	**3.5**	1
	FP	70	69.9	58	82.8	0.1	6.5	1
	MFCC-EUC	81.8	83.2	71.4	90.9	-0.4	6.1	0
	MFCC-G30	**86.9**	**89.1**	**73.7**	**92.9**	-1.1	6.2	0
Song D	**Users**	77.2	77.7	57.5	96.2	0	8.2	-
	MFCC-Var	60.8	60.1	57.2	71.1	1.4	**4.5**	**1**
	FP	57.3	54.6	50.9	73.4	1.3	7	1
	MFCC-EUC	66.1	65.1	59.7	84.6	1.9	7	1
	MFCC-G30	**77.8**	**83.0**	**63.0**	**86.6**	-0.5	8.8	0
Song E	**Users**	65.7	66.1	22.8	93.2	-0.4	15.4	-
	MFCC-Var	**67.8**	**68.5**	**56.2**	**82.3**	0.1	**8.1**	**0**
	FP	59.2	60.6	42.3	70.1	-0.5	9.8	0
	MFCC-EUC	34.1	34.2	11.6	62	0.1	15.4	1
	MFCC-G30	49.9	49.4	30.9	63.3	-0.4	9.7	1
Mean	**Users**	76.3	77.6	-	-	-	10.6	-
	MFCC-Var	69.4	70.3	-	-	-	**7.2**	-
	FP	**64.9**	64.7	-	-	-	8.4	-
	MFCC-EUC	64.0	64.2	-	-	-	9.0	-
	MFCC-G30	69.3	**71.6**	-	-	-	9.6	-

Table 1. Basic statistics of the distributions of the inter-users scores set and the users-automatic scores set. Values are in percent. Results of statistical test are shown too: $H = 0$ means that the two distributions are coherent, while $H = 1$ stands for a distributions mismatch. The codes for the automatic methods stand for: MFCC/Var = MFCCs clustered by standardized variogram, FP = fluctuation patterns, MFCC-EUC = MFCCs clustered by K-means, MFCC-G30 = MFCCs clustered by GMM/EM method. Best results in bold.

the test for song B. The method based on the clustering of the Mel coefficients by the variogram returns the highest scores for songs A and E. It also returns the second highest score for song B, although failing the test, but with the highest p-value (not shown in the table).

The best results are attained for the song A, where three of the four methods pass the test, while, for song B, none of them return a sufficient matching with the inter-users distribution. This last issue is basically related with the high agreement shown by the users (about 82%) that is hardly attained by the automatic methods. Quite the same situation occurs for song C and D, with high mean scores among the users lists (more than 84% and 77%, for songs C and D, respectively) approached by only two of the four methods proposed. Finally, the song E reveals a very low mean inter-users score (about 66%), well reflected by all the methods. Globally, all the methods show a good performance, with averaged mean values higher than 64%. The variogram-based approach shows the highest mean value, with 69.4%, very close to the result by the GMM/EM based method.

7. CONCLUSIONS AND FUTURE WORKS

A new approach based on the use of the standardized variogram for the clustering of the Mel coefficients for audio similarity evaluation has been proposed. The variogram is calculated on a reduced vector of ten lag elements and it is standardized by the global variance, in order to obtain comparable signature matrices for different songs. The method capability is evaluated on the base of a statistical comparison among the distributions of the matching scores computed among a set of users lists and the ones returned by the automatic method. Moreover, for a more complete assessment of the method performance, other three known methods employed in literature for audio similarity evaluation are computed, and their correspondent scores are compared.

Performances vary from quite poor to very good for all the methods, with mean matching scores varying from the lowest mean value of about 34% for the method based on the Euclidean distance and the highest value of about 87% for the method based on the clustering by GMM/EM. The averaged mean values reveal a good global performance of the method based on the variogram and on the GMM/EM approach, with quite poorer results by the other two ones. In practice, the variogram-based method proposed here works quite well and its performance can be compared with the one of other more popular methods that, in some cases, show a higher degree of computational complexity.

The capability of the method can be improved, by optimizing some calculation parameters, as the sampling of the distance lags values. Moreover, the theoretical variogram can be evaluated and its shape parameters can be taken into account to optimize the modeling of the spectral content of the song to improve the audio similarity assessment task. The evaluation task can be improved by increasing the number of users and broadening the test samples.

8. ACKNOWLEDGMENTS

This work was supported by the Ministerio de Educación y Ciencia of the Spanish Government, under Project No. TSI2007-61181, by the Ministerio de Industria, Turismo y Comercio of the Spanish Government, under Project No. TSI020501-2008-0117 and by the Junta de Andalucía, under Project No. P07-TIC-02783.

9. REFERENCES

[1] Jean-Julien Aucouturier and Francois Pachet. Music similarity measures: What's the use ? http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.9.5609, 2002.

[2] C. G. Dillon, C. Lloyd, and L. Philip. Identifying short-range and long-range structural components of a compacted soil: an integrated geostatistical and spectral approach. *Computers and Geosciences*, 29:1277–1290, December 2003.

[3] Stephen J. Downie. The music information retrieval evaluation exchange (mirex). http://www.dlib.org/dlib/december06/downie/12-downie.html.

[4] Jonathan T. Foote. Content-based retrieval of music and audio. In *Multimedia Storage and Archiving Systems II, Proc. of SPIE*, pages 138–147, 1997.

[5] John Haslett. On the sample variogram and the sample autocovariance for non-stationary time series. *The Statistician*, 46(4):475–485, 1997.

[6] Edward H. Isaaks and Mohan R. Srivastava. *An Introduction to Applied Geostatistics*. Oxford University Press, USA, January 1990.

[7] A. Kacha, F. Grenez, J. Schoentgen, and K. Benmahammed. Dysphonic speech analysis using generalized variogram. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, volume 1, pages 917–920, 2005.

[8] Davit Khachatryan and Sren Bisgaard. Some results on the variogram in time series analysis. *Quality and Reliability Engineering International*, March 2009.

[9] Daniel G. Krige. 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.

[10] Beth Logan. Mel frequency cepstral coefficients for music modeling. In *In International Symposium on Music Information Retrieval*, 2000.

[11] Beth Logan and Ariel Salomon. A content-based music similarity function. Technical report, Processing Languages – Document Style Semantics and Specification Language (DSSSL). Ref. No. ISO/IEC 10179:1996(E), 2001.

[12] George Matheron. The theory of regionalized variables and its applications. *Les cahiers du CMM de Fontainebleau*, 5, 1971.

[13] E. Pampalk. *Computational Models of Music Similarity and their Application to Music Information Retrieval*. PhD thesis, Vienna University of Technology, Vienna, March 2006.

[14] S. Sammartino. *Geostatistical models for environmental datasets*. PhD thesis, University of Napoli "Federico II", Napoli (Italy), March 2006.

[15] M. A. Stephens. Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737, 1974.

[16] Hans Wackernagel. *Multivariate Geostatistics: An Introduction With Applications*. Springer-Verlag Telos, January 1999.