BENFORD'S LAW FOR MUSIC ANALYSIS

Isabel Barbancho[1] **Lorenzo J. Tardón**[1] **Ana M. Barbancho**[1]
Mateu Sbert[2]
[1] Universidad de Málaga, ATIC Research Group, ETSI Telecomunicación,
Dpt. Ingeniería de Comunicaciones, Campus Teatinos, 29071 Málaga, Spain
[2] University of Girona, Institute of Informatics and Applications,
Campus Montilivi, 17003 Girona, Spain
ibp@ic.uma.es, lorenzo@ic.uma.es, abp@ic.uma.es, mateusbert@mac.com

ABSTRACT

Benford's law defines a peculiar distribution of the leading digits of a set of numbers. The behavior is logarithmic, with the leading digit 1 reflecting largest probability of occurrence and the remaining ones showing decreasing probabilities of appearance following a logarithmic trend. Many discussions have been carried out about the application of Benford's law to many different fields. In this paper, a novel exploitation of Benford's law for the analysis of audio signals is proposed. Three new audio features based on the evaluation of the degree of agreement of a certain audio dataset to Benford's law are presented. These new proposed features are succesfully tested in two concrete audio tasks: the detection of artificially assembled chords and the estimation of the quality of the MIDI conversions.

1. INTRODUCTION

Benford's law, also known as the 'first-digit law', describes a peculiar distribution of the leading digits of datasets of numbers, especially those related to the measure of 'real-life phenomena'. Unlike the *central limit theorem*, Benford's law states that the typical distribution of the leading digits of a large number of datasets, derived from the measure of several common variables follows a logarithm-shaped law.

Most of the measures from real-life (tax returns, street addresses, population number or length of rivers) seem to present this peculiar distribution. Many works have been published on Benford's law, mixing the empirical evidence with some more mathematical formalism.

Benford's law has been widely proposed as a discriminating tool for 'naturally-shaped' datasets [6] and even employed [8] or criticized [5] as a somewhat reliable diagnostic tool to detect a large variety of frauds.

In this paper, Benford's law is evaluated as a discriminator for audio signals. In particular it is employed to detect differences between natural and artificially created chords and real music and MIDI-generated music.

The article is organized as follows: in Section 2, Benford's law is discussed and its probabilistic framework is detailed. In Section 3, the three new audio features based on the evaluation of the degree of agreement of a certain dataset to Bendford's law are defined. These descriptors are widely employed in Section 4 for the aforementioned tasks, as part of the audio signals analysis. Finally, in Section 5, some conclusions are drawn.

2. BENFORD'S LAW

Benford's law affirms that the frequency of occurrence of the leading significative digit of a large dataset coming from real-life measurements, presents a peculiar histogram in which the height of the bars follows a logarithmic scale (see Figure 1).

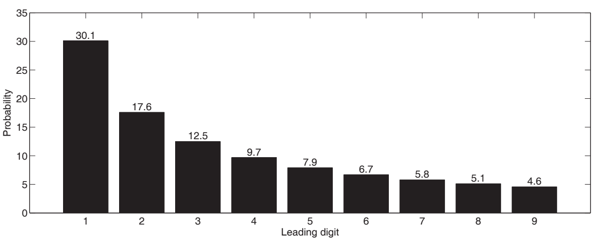


Figure 1. The logarithm-shaped distribution of the leading digits, following Benford's law.

More specifically, the probability value of the d-th digit is computed as follows:

$$P(d) = log_{10}\left(1 + \frac{1}{d}\right) \quad (1)$$

where d is the digit number.

Simon Newcomb [11] first described this peculiar behavior after the observation of the pages of the tables of common logarithms. He noticed that the logarithms beginning with the digit 1 were more frequently browsed than the others. In his two-page paper, he briefly described the empirical evidence of such observation, extending it to all the digits.

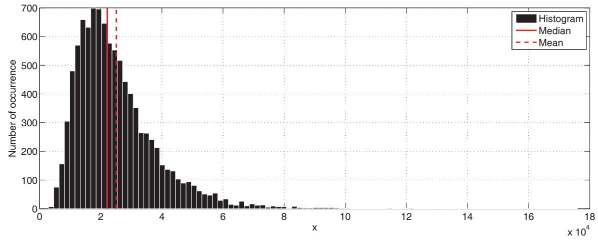
However, his work remained unknown for several years. In 1938, the *General Electric* physicist Frank Benford, apparently unaware of Newcomb's paper, formalized the same observations with a more consistent article published by the American Philosophical Society [4]. He included the formalization of the same law and a large amount of observations of real-life phenomena gathered during several years of research.

The rigorous mathematical discussion of the law was tackled several years after, and it is currently a matter of question. In 1976, the mathematician Ralph Raimi wrote about the mathematical explanation of the law, citing the 'scale-invariance' as one of the possible keys for interpretation of the phenomenon [14]. Theodore Hill [7], in 1995, described the statistical derivation of the law, while in 1997, Stephen Smith [15], in his book "The Scientist and Engineer's Guide to Digital Signal Processing" presented a rigorous complete description, under the point of view of the signal processing.

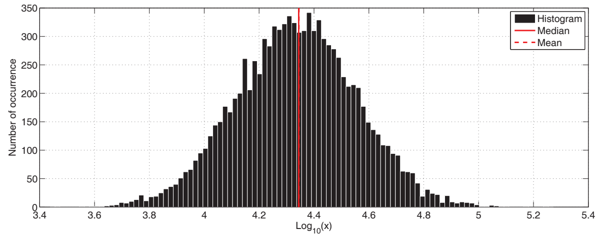
Nowadays, Benford's law is a well defined probabilistic problem, and it has been demonstrated that it is based on an intrinsic property of a large number of real-life datasets. According to the central limit theorem [13], the distribution of a certain measure of a quantity follows a normal distribution. The larger the amount of data, the closer the fit of the sample histogram to the Gaussian distribution. Nevertheless, when a single measure is iteratively repeated, its variance tends to be steady and to robustly define the range of variability of the quantity measured. Usually, it limits the width of the distribution to few orders of magnitude. In fact, it is infrequent that a series of iterative measures of the same variable could span across a wide range of values.

Also, if we multiply groups of random numbers, each following a normal distribution, we will obtain a new dataset following the so called 'log-normal' distribution [9]. Its name derives from the dome-shaped histogram that this kind of distribution shows, when it is represented on a logarithmic scale. In log-normal distributions, 95% of the values are distributed within the mean μ minus twice the standard deviation σ and the mean μ plus twice the standard deviation σ, on the logarithmic scale. This leads to an accumulation of values on the left edge of the distribution, on the linear scale [15]. Actually, in log-normal distributions the median is lower than the mean and they present large positive values of skewness [9] (see Figure 2).

The fact that the log-normal distribution usually derives from the combination of normally distributed variables, leads one to assume that, in nature, it is as common as the normal distribution [15]. Most of the datasets of real-life variables are log-normally distributed, especially those with only-positive values, where the intrinsic limitation leads to an increase of probability around the smallest values. Most of these datasets follow Benford's law. In environmental pollutants datasets, for instance, most of the measures are typically very low and only few of them are larger than their mean. Moreover, these variables are typically only positive, but they usually show very low values, very close to zero. This leads to a compression of the



(a) The histogram of a log-normal shaped dataset (linear axis).



(b) The histogram of a log-normal shaped dataset (logarithmic axis).

Figure 2. An example of log-normal shaped dataset in linear and logarithmic axis. The median and the mean are represented with continuous and dashed line, respectively. Note that the median and the mean coincide when the histogram is spaced on the logarithmic axis (the distribution is normally-shaped). The histogram bins have been equally spaced on the logarithmic axis, such to define a constant width of the bars.

histogram toward the minimum, resulting in a typical log-normal distribution.

Nevertheless, the shape of the histogram is not sufficient to be an index of the degree of fit to Benford's law. Usually, the log-normal distributions derived from the combination of multiple normal distributions (with different widths) are broader than them, because of the larger range of variability they present. In fact, it is the width of these kinds of distributions, that is key to understand their relation to Benford's law. Smith [15] shows how the degree of fit to the law of a certain dataset is a mere question of distribution width. The broader the distribution of the data, the more accurate the fit to the theoretical law.

This is a very important issue, related to the data manipulation by humans. The most common way to systematically extract the leading digit of a number is to multiply or divide it by ten, until it reaches a value between 1 and 9.9 periodic. In particular, the number must be divided by 10, if it is higher than 10 and multiplied by 10, if it is lower than 1.

Thus, for instance, the number 0.00567 will be multiplied by 10 three times to obtain the number 5.67, whose integer part (5) is taken into account as the leading digit. Similarly, for the number 7865, it has to be divided by 10 three times to obtain 7.865, and the correspondent leading digit (7).

This 'human-driven' mechanism is primarily responsible for dependence of the distribution of the leading digits on the logarithmic law [15]. Hence, the amount of dependence, namely the degree of fit to Benford's law, depends on the broadness of the original data distribution. If the data span across a large number of orders of magnitudes, with respect to unity, they will need several steps of multiplications/divisions to be scaled to range between 1 and 9.9 periodic. Conversely, if the dataset ranges from 1 to 9, the numbers will not require any operation. The impact of these kinds of manipulations is directly related to the degree of agreement to Benford's law.

3. BENFORD'S LAW BASED AUDIO FEATURES

In order to evaluate the degree of agreement of a certain dataset to Benford's law, several approaches can be employed. The task is to obtain new features to be used as comparative measure among the different audio elements to be classified. In this section three new features will be extracted: the one-scaling-test, the Fourier-based method and the goodness-of-fit test.

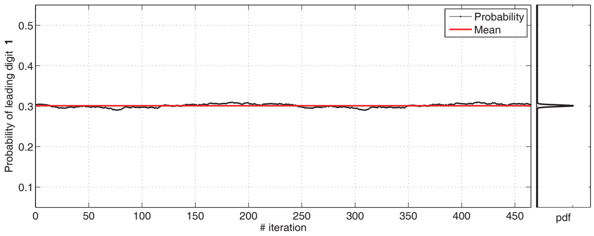
3.1 The one-scaling-test

Raimi [14], still speaking about a "*universal law*", introduced the scale-invariance principle to define the validity of the law. He affirmed that "*...since God is not known to favor either the metric system or the English system...*", Benford's law must be scale-invariant. Smith [15] formalized a test based on the scale-invariance of the law, by measuring the variation of the probability of occurrence of the leading digit 1, when the dataset is iteratively multiplied by a constant.

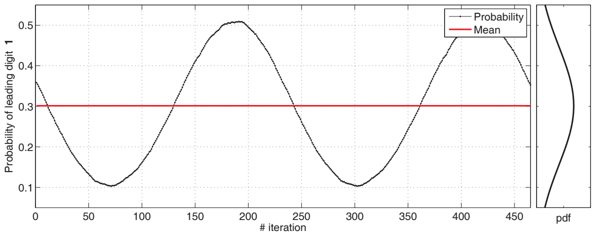
The theoretical probability of occurrence, by Benford's law, of the first leading digit is 0.301. If the empirical probability of the digit 1 of a certain dataset is close to this value, we can suppose that the dataset follows (potentially) Benford's law. This is obviously not sufficient. Recalling the concept about the scale-invariance, we can affirm that if the dataset follows the law, the empirical probability of occurrence of its leading digits (the digit 1 in this case) should not vary, or vary very weakly, if the dataset is iteratively multiplied or divided. The so called *one-scaling-test* proposed by Smith [15] exploits this property to evaluate the agreement of a certain dataset to Benford's law.

If we take two log-normally distributed datasets with equal mean, 10, and different standard deviation, 0.5 and 3, respectively, and we multiply them iteratively by a constant (e.g.: 1.01), we will observe a certain variation of the probability of occurrence of the first leading digit around the value 0.301 (Figure 3).

A broader distribution presents a much weaker variation of the probability of occurrence of the first leading digit around the value 0.301, than a narrower distribution. The means of the distribution of the probability values are 0.3010 and 0.3013, for the broader and the narrower distribution, respectively. That is, they both follow Benford's law, showing a value close to the expected theoretical probability (0.301). However, their standard deviations 0.0069 and 0.1450 reveal a much larger variation around the mean



(a) The variability of the probability of occurrence of the leading digit 1 for a broad log-normally distributed dataset ($\sigma = 3$).



(b) The variability of the probability of occurrence of the leading digit 1 for a narrow log-normally distributed dataset ($\sigma = 0.5$).

Figure 3. An example of the effect of the width of the (log-normal) distribution on the one-scaling-test. Means are represented with thick lighter line. The equivalent PDFs are displayed on the right side of the plots.

for the dataset with the narrower distribution. Although both datasets seem to follow Benford's law, the broader one required a heavier manipulation of the original data to extract the leading digits and it emphasized the logarithmic pattern attributed to their distribution, leading it to approach the theoretical law closer.

Note that in both cases, the variation of the probability shows a periodic pattern due to the factor chosen for the multiplication. The leading digit is unchanged when the numbers are multiplied by 10. In our example, this occurs every 232 times ($1.01^{232} \approx 10$).

The one-scaling-test presents a main drawback related to the high computational cost derived from the iterative multiplication of the whole dataset. If we consider one single minute of an audio signal recorded at a sampling frequency of 44.1 kHz, we have to handle with a vector of more than 2.5 millions samples. If we want to multiply this dataset at least $232 + 1$ times (to observe at least one whole period), we must do more than 600 millions of operations. In the case of exploiting Benford's law in a classifier tool for music genres, we should have to handle hundreds of songs, each of them with a length of several minutes. This would become an unfeasible task from the point of view of the computational cost.

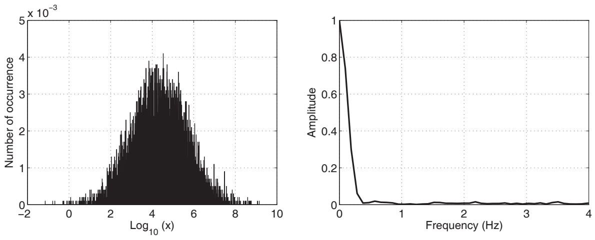
3.2 The Fourier transform-based method

Smith [15] reinterprets the problem from the point of view of signal processing. He proves that the degree of agreement of a certain dataset to Benford's law, can be estimated by evaluating the behavior of the Fourier transform (FT) of the normalized histogram in logarithmic axes. In particular, the measure of how fast the transform falls, from its maximum value (1 at frequency 0) to its minimum value (zero at some frequency higher than zero), is directly related to the width of the distribution measured with the normalized histogram and, consequently, with the degree of correspondence with the law.

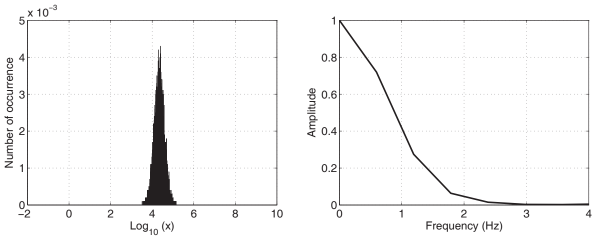
Ideally, in order to follow perfectly Benford's law, the Fourier transform should present a unitary value at frequency zero and a zero value at all the remaining frequencies. This would occur if the distribution was uniform from $-\infty$ to $+\infty$ [12].

In real-life, this does not occur. Hence, the faster the Fourier transform drops to zero, the closer the agreement of the dataset to Benford's law. In particular, Smith defines the value at frequency 1 as the threshold to discriminate between the agreement or not of the dataset to the law [15]. If the transform of the histogram in logarithmic axes (denoted as PDF) falls to zero before frequency $1Hz$, the correspondent dataset follows Benford's law. If it does not occur, the dataset does not follow the law. In practice, the value of the PDF at $f = 1Hz$ is a reliable index of the degree of agreement with the law.

In Figure 4, an example of the application of the Fourier transform to the histograms of the dataset tested in the previous section, is shown.



(a) Broad log-normally distributed dataset. Left: distribution on a logarithmic axis. Right: Fourier transform of the distribution (PDF).



(b) Narrow log-normally distributed dataset. Left: distribution on a logarithmic axis. Right: Fourier transform of the distribution (PDF).

Figure 4. Example of the application of the Fourier transform for the estimation of the agreement of the data to Benford's law. The transform of the broader distribution drops to zero faster than the narrower one, revealing a closer correlation with the law.

The distribution of the data shown in Figure 4(a) is broader than the one in Figure 4(b). Actually the two datasets are the same that were previously analyzed in Figure 3, with standard deviation 3 and 0.5, respectively. The PDF of the broader distribution falls to zero much faster than the narrower one. In particular, the amplitude of the PDF at frequency $1Hz$ is 0.0023 and 0.4184, for the broader and the narrower distribution, respectively. This issue reveals a closer agreement to Benford's law of the broader distribution, as observed previously.

Note that unlike the one-scaling-test, the method based on the Fourier transform has a reasonable computational cost. Furthermore, this method returns a higher discriminant range for the two datasets: The ratio between the two standard deviations of the one-scaling-test is about 20, while the ratio between the two values of the transforms at frequency $1Hz$ is about 180. If the aim of the application of Benford's law is a boolean discrimination of the data, then the Fourier transform-based method is efficient.

3.3 The χ^2 divergence and the goodness-of-fit test

An alternative to the the two empirical methods proposed so far, is the well known χ^2 test [9]. It is called the *goodness-of-fit* test and it returns a measure of how well an empirical distribution fits a theoretical one.

The divergence is calculated as follows:

$$D = \frac{(f(d) - P(d))^2}{P(d)} \tag{2}$$

where $f(d)$ is the empirical relative frequency of the digit d and $P(d)$ stands for its theoretical probability defined, in our case, by Benford's law, detailed in equation (1).

The null hypothesis $H0$ is verified if its associated probability (the p-value) does not exceed the significance level fixed a priori. This probability value, when the test is passed, can be employed as additional information for the measure of the agreement to Benford's law.

The goodness-of-fit test applied to the two datasets analyzed before, returns a divergence value of 0.2484 and 0.0009, for the narrower and the broader distribution, respectively. Actually, the narrower distributed dataset did not pass the test. In Figure 5, the two empirical distributions of the leading digits are shown.

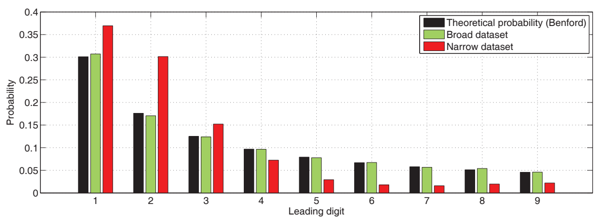


Figure 5. The distribution of the leading digits for the two datasets of the previous example. Both empirical distributions are compared against the theoretical values.

Once again, the broader distribution reveals a larger correlation with Benford's law, than the narrower one. Note that the ratio between the two divergences (about 280) is even larger than the one measured between the two values of the transform in the previous example.

Nevertheless, the approach based on the Fourier transform does not need to extract the samples in the dataset and it is, therefore, more efficient.

4. EVALUATION OF BENFORD'S LAW PERFORMANCE IN PATTERN RECOGNITION TASKS FOR MUSIC SIGNALS

In this section, the performance of the proposed features based on quantitative measurements of agreement to Benford's law is evaluated in two concrete audio tasks.

4.1 Real and artificially assembled chords

Often, the methods employed in the evaluation of the algorithms for multi-pitch estimation are based on the usage of ground-truth datasets of artificially assembled chords, i.e. made up by the summation of individual waveforms of the single notes that compose the chords. In this context, this procedure leads to cleaner spectra that can be more easily analysed. Benford's law based audio features are employed to discriminate between real and artificially assembled chords. A set of 230 chords has been examined. The half of them are real chord [3] and the other half are the same chords but artificially assembled adding single notes. The two sets of chords do not reveal any kind of significant difference when submitted to a perceptual evaluation. They sound practically the same.

Using this data set, the descriptors to evaluate the agreement of the data to Benford's law have been calculated for each pair of chords (real and artificially assembled). The ones-scaling test has not been performed because of its high computational costs. In order to evaluate the performance of the new Bendford's law based features, they have been compared against a set of time and frequency features commonly used for the music classification task (RMS, ZCR, CER, SPF) [16].

The descriptors defined in this context reflect a notable discrepancy between the two classes of chords. Surprisingly, an average value of the 30% of the samples (12 out of 115 for the artificial chords and 56 out of 115 of the real chords) did not pass the χ^2 test. The signals showed rather skewed distributions on the logarithmic axis with the consequent decrease of the level of agreement to Benford's law.

A knn classifier has been adopted here to perform the classification of the chords using both the set of single features selected and the two groups of features with and without the two Benford's law-based descriptors. As it is shown in Table 1, Benford's law-based features behave rather well when compared against the typical features for audio classification. Also, the multidimensional set of descriptors improves its performance with the inclusion of the two Benford's law-based features. It is interesting to note that the artificial chords returned smaller values of $PDF(f = 1\text{Hz})$ than the real chords (see Figure 6).

4.2 Quality of MIDI conversion

Recalling the 'Nature-dependence' of Benford's law, we formulate the hypothesis that the agreement of MIDI [10] audio to Benford's law, could be used as a ranking measure for the quality of automatic MIDI converters. Two software tools for automatic MIDI conversion were tested: the freeware software AMAZING MIDI (v1.70) by arakisoftware [2] and the shareware software AKoff Music Composer (v2.0) by the AKoff Sound Labs [1]. Both of them present at least two different configuration sets. In particular, the AKoff software has been run with and without the application of the 'overtones filtering', a utility to filter the highest harmonics of the spectrum, while the AMAZING MIDI software has been executed with and without a time and an amplitude filter (to reduce the range of amplitude and note duration).

The term "quality of a MIDI conversion" is a rather subjective concept, i.e., it may depend on the person who is evaluating that quality. Therefore, the sounds of the automatic conversion tools have been listened carefully by a team of ten expert musicians who have evaluated personally both the similarity between the converted track and the original one, and the overall quality of the MIDI audio. Each listener had to rank the MIDI converters with a score in the range 0 (the worst quality) to 100 (the best quality). Table 2 shows the mean of the subjective test scores obtained by each tool/configuration.

In Figure 7, an example of the test performed, applied to the song 'Come sei veramente' by the pianist G. Allevi, is shown. The original track returned the smallest value in the ones-scaling test, $PDF(f = 1\text{Hz})$ and the χ^2 diver-

Feature	Success rate (%)
Benford's law-based features	
$PDF(f = 1\text{Hz})$	80.87
χ^2 divergence	71.74
Time and Frequency features	
Root mean square	82.61
Zero crossing rate	68.70
Cepstrum residuals	58.26
Spectral flux	72.61
Grouped-feature set	
Time and Frequency features set	79.57
Benford's law-based features added	82.17

Table 1. Real and artificial chord classification accuracy of the single-feature tests and the grouped-feature tests.

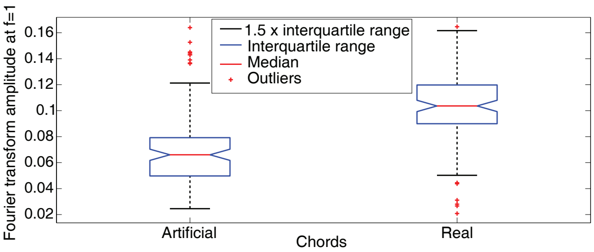


Figure 6. Box-whiskers plot of the amplitude of $PDF(f = 1\text{Hz})$, for artificial and real chords. The non-overlapping notches, indicating the 95% confidence interval of the two medians, reveal good discrimination power of the analyzed feature.

Software/configuration	Mean score
AKoff with overtone control	27/100
AKoff without overtone control	48/100
AmazingMIDI with filters	75/100
AmazingMIDI without filters	80/100

Table 2. Mean subjective ranks of the four combination of tool and configuration employed in the MIDI-quality test.

gence, with respect to the other four MIDI versions. The two outcomes of the AKoff software returned the largest values of each descriptor, revealing the lowest accordance to Benford's law. Note the relation between features extracted and the subjective ranks in Table 2. Therefore, the accordance to Benford's law provide us with a measure of the quality of the MIDI converters.

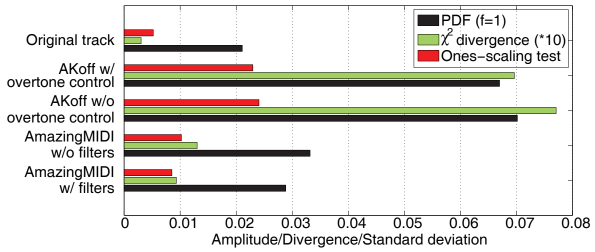


Figure 7. The three Benford's law-based features calculated for the track 'Come sei veramente' by the Italian pianist Giovanni Allevi. Divergence values are multiplied by 10 for displaying purposes.

5. CONCLUSIONS

In this paper, it has been shown how Benford's law can be conveniently exploited to extract useful features that can be successfully used in different audio pattern recognition tasks. Three new Benford's law based audio features based on different measurement of the agreement to Benford's law have been proposed.

Two concrete tasks have been addressed to highlight this novel context of application of Bendford's law for audio signal. For chord analysis, the new proposed features are rather compelling as good discriminators when compared against other typical features for speech and audio classification and also the results obtained for the determination of the quality of the automatic MIDI conversions are promising.

Therefore, through this paper it has been illustrated how Benford's law, that substantially arises as a matter of shape and width of the distribution of the leading digits of the data, can be conveniently exploited for audio classification problems.

6. ACKNOWLEDGEMENT

This work has been funded by the Ministerio de Economía y Competitividad of the Spanish Government under Project No. TIN2013-47276-C6-2-R and Project No. TIN2013-47276-C6-1-R. This work has been partially done at Universidad de Málaga, Campus de Excelencia Internacional (CEI) Andalucía Tech.

7. REFERENCES

[1] AKoff-Sound-Labs. AKoof music composer, wav-to-midi converter. `http://www.akoff.com/music-composer.html`, 2011.

[2] Arakisoftware. AmazingMIDI, wav to midi converter for music transcription. `http://www.pluto.dti.ne.jp/~araki/amazingmidi`, 2011.

[3] Ana M. Barbancho, Isabel Barbancho, Lorenzo J. Tardón, and Emilio Molina. *Database of Piano Chords. An Engineering View of Harmony*. Springer-Verlag, New York, NY, USA, 2013.

[4] Frank Benford. The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78(4):551–572, 1938.

[5] Joseph Deckert, Mikhail Myagkov, and Peter C. Ordeshook. The irrelevance of benfords law for detecting fraud in elections. *Caltech/MIT Voting Technology Project*, 1(9), 2010.

[6] Esteve del Acebo and Mateu Sbert. Benford's law for natural and synthetic images. In *Computational Aesthetics*, pages 169–176. Eurographics Association, 2005.

[7] Theodore P. Hill. A statistical derivation of the significant-digit law. *Statistical Science*, 10(4):354–363, 1995.

[8] Brian A. Jacob and Steven D. Levitt. Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, 118(3):843–877, 2003.

[9] Averill M. Law and W. David Kelton. *Simulation Modeling and Analysis*. McGraw Hill, Inc., 1991.

[10] Robert A. Moog. Midi: Musical instrument digital interface. *Journal of the Audio Engineering Society*, 34(5):394–404, 1986.

[11] Simon Newcomb. Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, 4(1):39–40, 1881.

[12] Alan V. Oppenheim and Ronald W. Schafer. *Discrete-Time Signal Processing, 3/E*. Prentice Hall, 2010.

[13] Athanasios Papoulis and S. Unnikrishna Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw Hill, 2002.

[14] Ralph A. Raimi. The first digit problem. *The American Mathematical Monthly*, 83(7):521–538, 1976.

[15] Steven W. Smith. *The scientist and engineer's guide to digital signal processing*. California Technical Publishing, San Diego, CA, USA, 1997.

[16] Lorenzo J. Tardón, Simone Sammartino, and Isabel Barbancho. Design of an efficient music-speech discriminator. *The Journal of the Acoustical Society of America*, 127:271–279, 2010.