



## APLICAÇÃO DO MÉTODO DFA NA IDENTIFICAÇÃO DE CORRELAÇÕES E FRACTALIDADES EM SÉRIES TEMPORAIS SONORAS

Ivan Marcelo Laczkowski, UTFPR-CM,  
laczkowski@ig.com.br  
Mirian Sousdaleff Laczkowski,, UTFPR-CM,  
miriansousdaleff.87@gmail.com

### INTRODUÇÃO

Séries temporais são conjuntos de dados coletados, geralmente, em intervalos regulares de tempo. Elas distinguem-se de dados coletados aleatoriamente porque representam a evolução temporal da dinâmica de uma população ou processo. A ordenação linear dessas séries assumem lugar de destaque na análise de dados, com um conjunto especializado de técnicas para estudo (MORETTIN, 2006).

Nas últimas décadas, foi descoberto que muitos sistemas dinâmicos vivos e não vivos, quando investigados por meio de mecanismos multivariados apresentam correlações e propriedades multifractais. E, muitos dos registros existentes, estão organizados na forma de séries temporais (MORETTIN, 2006). Neste cenário, é muito importante identificar se numa série ou sequência de dados existe correlação, isto é, se há um efeito de memória de longa duração no decorrer do tempo.

Quando uma série apresenta correlação de memória longa, suas propriedades podem ser investigadas utilizando conceitos e métodos da teoria fractal (ADDISON, 1997). Esta teoria é empregada no estudo de flutuações no tempo ou objetos no espaço que possuem autossimilaridades e que não podem ser investigados por meio de uma única escala de medida absoluta. Em geral, fractais apresentam irregularidades no espaço ou no tempo, como as camadas de uma cebola. Partes de um objeto ou sequência fractal são cópias exatas ou estatísticas do todo e podem se tornar iguais ao objeto original por ampliação e deslocamento (ADDISON, 1997).

Um processo aleatório pode apresentar memória de longa duração quando um evento passado tem efeito sobre eventos futuros, que podem ser esquecidos à medida que o tempo avança. Assim, para definir processos com memória longa utiliza-se a função de autocorrelação  $C(s)$ .

Quando uma série temporal é correlacionada, um valor  $Y(t)$  no tempo  $t$  correlaciona com um valor  $Y(t+s)$ , em que  $s$  é um incremento qualquer de tempo no futuro. No caso mais trivial, em que a série não é correlacionada  $C(s) = 0$  para  $s > 0$ . Quando a série é correlacionada,  $C(s)$  é uma função decrescente para diferentes valores de  $s$ . Se a função  $C(s)$  decai exponencialmente, diz-se que a série possui correlação de curto alcance ou memória de curta duração. Por outro lado, num processo de memória longa, a função de autocorrelação decai como uma lei de potência com  $C(s) = Cs^{-a}$ . Em que



$C$  é uma constante,  $s$  é o tempo entre as observações e  $a$  representa o expoente de correlação, assumindo valores entre 0 e 1 (TAQQU, 1995).

Diversas técnicas foram propostas para identificar correlações de longo alcance em séries temporais (MANDELBROT, 1983; ADDISON, 1997). O método, amplamente utilizado para esse estudo é o método *Detrended Fluctuation Analysis* – DFA (PENG et al. 1994). Por meio dessa metodologia, busca-se estimar a média das flutuações  $F'(s)$  na série em escalas de tempo variável de tamanho  $s$ , e verificar como  $F'(s)$  depende da escala  $s$ . A função de flutuação  $F'(s)$  deverá apresentar um comportamento do tipo lei de potência para diferentes valores de  $s$ , isto é,  $F'(s) = s^h$ . Em que  $h$  é o expoente de Hurst. A variação temporal do expoente de Hurst para uma dada série indica a existência de flutuações não estacionárias, indicando um processo multifractal (PENG, 1994).

Atualmente, a análise multifractal tem sido aplicada com muito sucesso em diversas áreas, tais como séries temporais de indicadores econômico (KANTELHARDT, 2002), meteorologia (SHUKLA, 1998; SUTTON, 1997), ciências da vida (PENG, 1994) dentre outras áreas.

Considerando o exposto acima, o objetivo deste trabalho é apresentar a aplicação do método DFA em séries temporais sonoras, e por meio da análise do expoente de Hurst, identificar correlações de longo tempo e fractalidades.

## MATERIAIS E MÉTODOS

No presente trabalho foram selecionadas três séries temporais para aplicação do método DFA. As séries estudadas foram obtidas de gravações provenientes da internet e recortadas por meio de um *software* para edição de músicas<sup>1</sup>. Uma das séries representa um sinal sonoro do tipo ruído branco<sup>2</sup>; e as outras duas, são sons musicais de gêneros diferentes. Os sons musicais escolhidos para análise foram a canção Lutar Pelo Que é Meu (música do gênero *rock*) da banda Charlie Brown Jr<sup>3</sup> e o Concerto No. 1 in E (La Primavera) de Antonio Vivaldi<sup>4</sup> (música do gênero clássico).

### O método DFA

Como mencionado na seção anterior, a metodologia de análise de flutuações destendenciadas (*Detrended Fluctuations Analyses* - DFA) é um método amplamente utilizado pelos pesquisadores na investigação de escalas fractais e identificação de correlações de longo alcance em séries temporais.

<sup>1</sup> <http://audacity.sourceforge.net/?lang=pt>

<sup>2</sup> <http://www.youtube.com/watch?v=GRcppUVFoso>

<sup>3</sup> <https://www.youtube.com/watch?v=cHVGPFGBes>

<sup>4</sup> <http://www.youtube.com/watch?v=O6NRLYUThrY>



Por meio deste método, evita-se a detecção de falsas correlações que são de origem não estacionárias nas séries estudadas (PENG, 1994). E, neste contexto, o método tem por objetivo eliminar as tendências determinísticas da série original e estudar os dados de uma forma destendenciada.

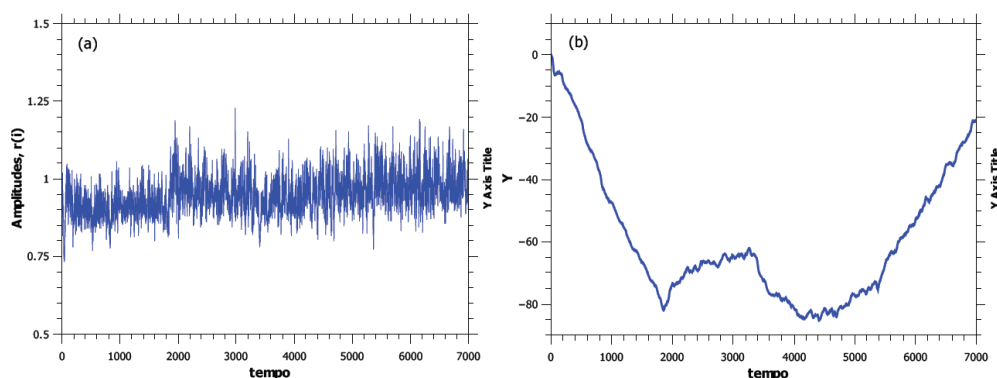
Para realizar uma análise DFA, quatro passos devem ser seguidos (BULDYREV, 1995; PENG, 1994): i) a partir da série original obtém-se a série acumulada; ii) dividir a série acumulada em janelas ou subconjuntos de tamanho  $s$ ; iii) remover a tendência local em cada janela por meio de um ajuste polinomial; e iv) calcular a função de flutuação  $F'(s)$  em termos do tamanho das janelas.

Para ilustrar de uma maneira mais formal os procedimentos indicados acima, considere uma série temporal, conforme ilustrado na figura 1 (a), cujos elementos são representados por  $r(i)$ , com  $i = 1, 2, 3 \dots N$ . A série acumulada será obtida por meio da equação 1,

$$Y(i) = \sum_{k=1}^i [r(k) - \langle r \rangle] \quad (1)$$

Em que,  $Y(i)$  é a série acumulada ou integrada (ver figura 1 (b)), e  $\langle r \rangle$  é o valor médio de  $r(i)$ , obtida por,

$$\langle r \rangle = \frac{1}{N} \sum_{i=1}^N r(i) \quad (2)$$



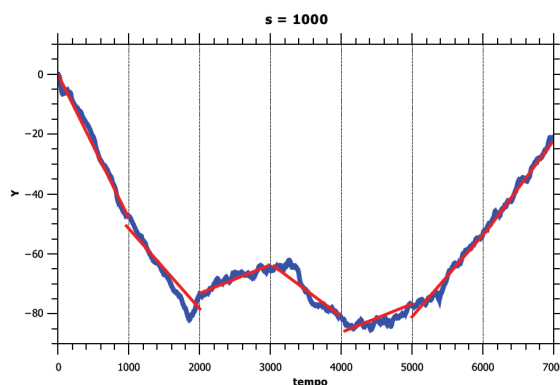
**Figura 01** – (a) Série temporal original. (b) série acumulada após ter sido retirada a média.

Em seguida, a série acumulada  $Y(i)$  é subdividida em janelas ou intervalos de tempos iguais e não sobrepostos de tamanho  $s$  e índice  $n$ , em que  $n = 0, 1, 2, 3 \dots Ns-1$  com  $Ns$  sendo a parte inteira de  $N/s$ . Nas janelas identificadas pelo índice  $n$ , ajusta-se os dados por um polinômio de tendência local de ordem  $l$ . Na literatura, é muito comum os termos DFA-1, se  $l = 1$ , DFA-2, se  $l = 2$  e dessa forma segue. Em seguida, a tendência local da série acumulada deve ser subtraída conforme indicado na equação 3.



$$W_s(i) = Y(i) - Y_n(i) \quad (3)$$

Em que,  $W_s(i)$  é a função das diferenças,  $Y_n$  é um polinômio ajustados aos dados na janela de tamanho  $s$  e índice  $n$ . A figura 2 ilustra esse procedimento.



**Figura 02** – Ajustes lineares para  $s = 1000$ .

Neste momento, é muito importante ressaltar que se o tamanho da janela, representada por  $s$ , for muito grande, próximo do comprimento da série, um número pequeno de janelas serão obtidas, por outro lado, se o tamanho da janela for pequeno, poucos pontos serão ajustados. Como esses extremos não são interessantes, define-se que o tamanho máximo de uma janela não seja maior que um décimo do comprimento da série; e que a menor janela seja maior que seis amostras (BULDYREV, 1995).

No último passo, calcula-se a variância de  $F'(n)$ ,

$$F_s^2(n) = \frac{1}{s} \sum_{i=1}^s W_s^2 [(n-1)s + i] \quad (4)$$

Tomando uma média sobre as amostras de  $W_s^2(i)$  na janela de índice  $n$ , define-se a função de flutuação  $F'(s)$  dada por

$$F'(s) = [(F_s^2(n))]^{\frac{1}{2}} \quad (5)$$

Em que  $\langle F_s^2(n) \rangle$  é a média de  $F_s^2(n)$  calculada sobre todas as  $Ns$  janelas, definida como

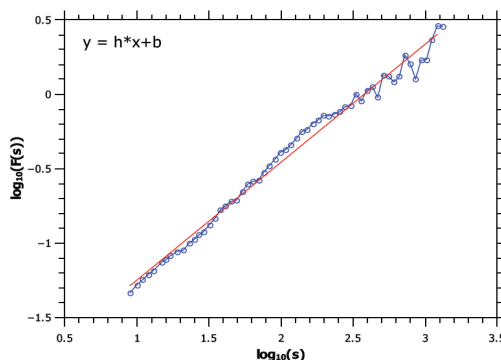
$$\langle F_s^2(n) \rangle = \frac{1}{Ns} \sum_{n=1}^{Ns} F_s^2(n) \quad (6)$$

Como a variância aumenta com o tamanho das janelas, em geral, a função de flutuação  $F'(s)$  será uma função crescente. Se o crescimento for do tipo lei de potência, conforme a equação 7,

$$F'(s) = s^h \quad (7)$$



Indica que a série possui fractalidade ou auto-similaridade. E, assim, o expoente de Hurst  $h$  poderá ser obtido por meio de um ajuste linear ( $y = hx + b$ ) de um gráfico de  $\log F'(s)$  por  $\log s$ , como indicado na figura 3.



**Figura 03** – Ajuste linear para obtenção do expoente de Hurst  $h$ .

É neste cenário que o expoente de Hurst revela propriedades de correlação na série em estudo. Alguns casos são muito importantes: i) se a série não for correlacionada (ruído branco), o valor do expoente  $h$  deve ser igual a  $0,5$ ; ii) se a série possuir correlações de longa duração, obedecendo uma lei de potência, com expoente de correlação  $a$  entre  $0$  e  $1$ , o expoente  $h$  deverá estar entre  $0,5$  e  $1$ . Neste caso, o expoente  $h$  é relacionado com  $a$  pela relação  $h = 1 - a/2$ , e a série é dita persistente. Se a série em estudo apresentar anticorrelações de longa duração, o expoente  $h$  deverá assumir valores entre  $0$  e  $0,5$ , e a série é dita anti-persistente; iii) se  $h = 1$ , a série representa um ruído  $1/f$ , se  $h > 1$ , haverá correlações, porém não na forma de lei de potência, e se  $h = 1,5$  a série representa um ruído browniano

Além disso, se a análise DFA for realizada numa versão embaralhada da série original, espera-se obter  $h = 0,5$ , pois no processo de embaralhar a série original, todas as possíveis correlações serão perdidas (KANTELHARDT, 2001).

## RESULTADOS E DISCUSSÕES

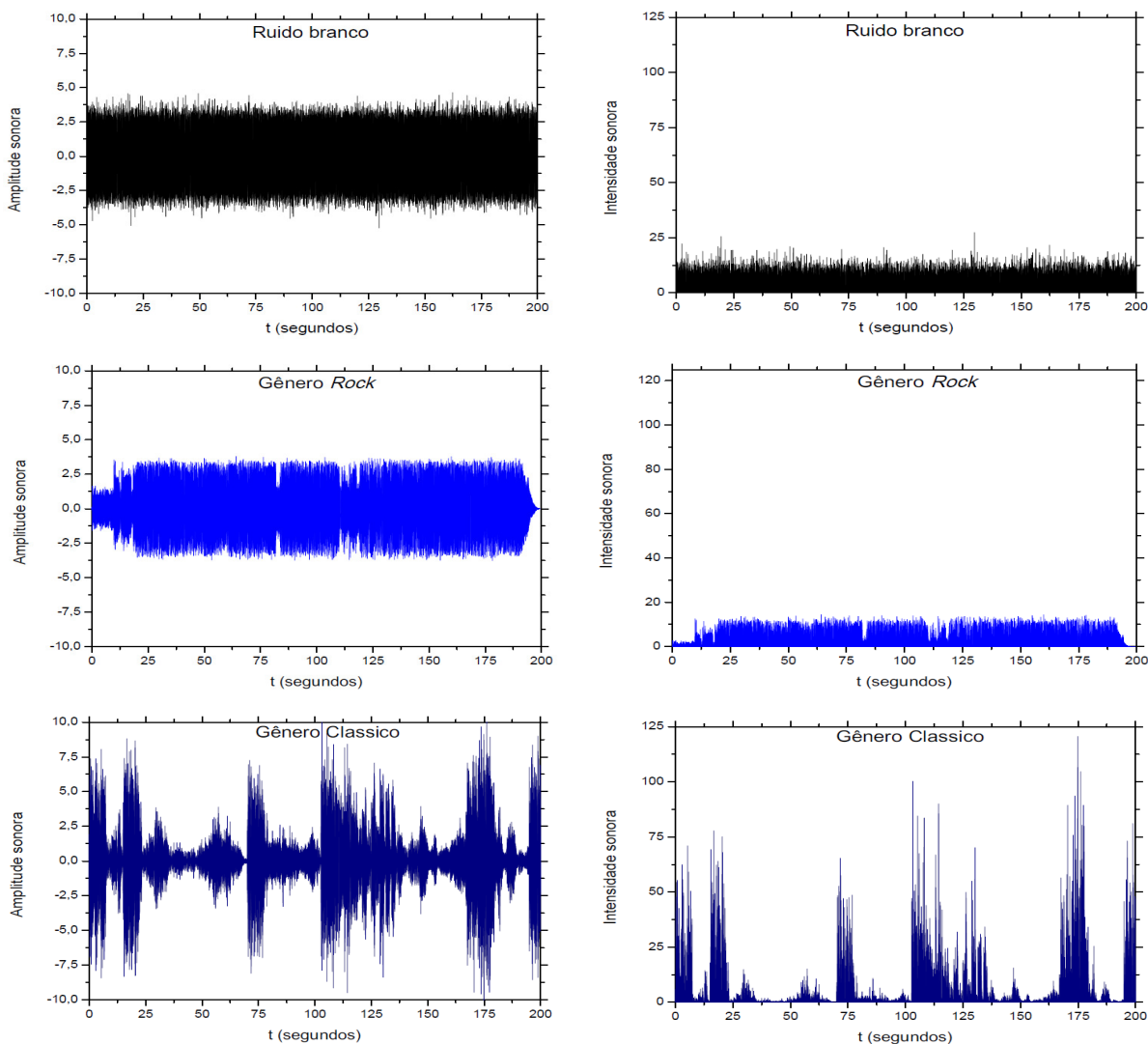
Para possibilitar a análise das séries temporais, o algoritmo *DFA* descrito acima foi inserido no computador por meio do *software* Matlab<sup>5</sup>. E, para realização dos cálculos necessários, as séries foram normalizadas de modo que elas apresentem médias zero e desvio padrão unitário. Com as séries normalizadas, é muito importante obter as séries das intensidades sonoras, pois estas últimas estão diretamente ligadas à quantidade de energia que um dado som possui. As séries das intensidades sonoras são obtidas quando se eleva ao quadrado a série das amplitudes sonoras (HALLIDAY,

<sup>5</sup> <http://www.mathworks.com/products/matlab/>





2012). A figura 4 mostra as séries das amplitudes (na esquerda) e as séries das intensidades (na direita).



**Figura 04** – Amplitudes sonoras normalizadas (na esquerda) e intensidades sonoras (na direita). De cima para baixo: Séries de ruído branco, Séries do gênero rock e Séries do gênero clássico.

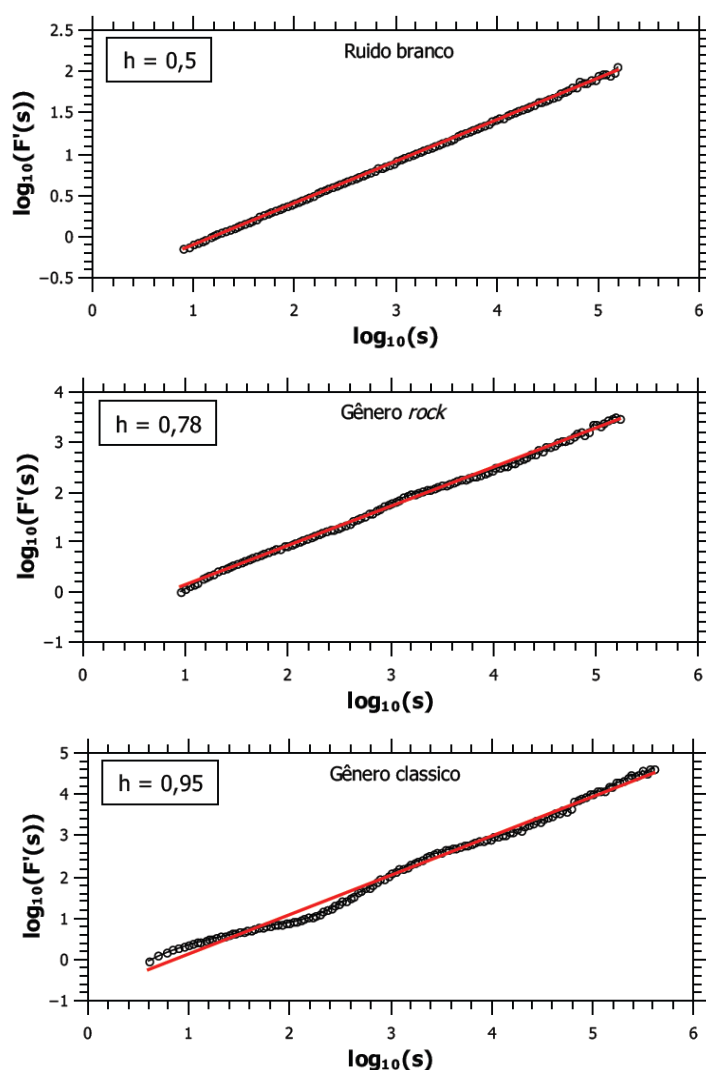
Nas séries da figura 4, observam-se similaridades e diferenças. Quando as séries das amplitudes ou intensidades da música do gênero rock são comparadas com as séries dos ruídos percebe-se que ambas apresentam similaridades. Suas amplitudes e intensidades são uniformes ao longo do tempo. Por outro lado, as séries das amplitudes e intensidades da música clássica quando comparadas com as séries de ruídos e com as séries da música do gênero rock observa-se uma sequência de trechos de valores altos e pequenos.

É natural esperar que uma série musical seja uma sequência de dados correlacionados, e que não deveria apresentar similaridades com um ruído branco, que são dados sem qualquer correlação. E,



neste contexto, o método DFA pode ser utilizado para indicar a existência de correlações em uma série temporal. No caso de uma série musical, o DFA poderá dar informações quantitativas que possibilita a distinção entre músicas mais e menos elaboradas conforme o valor do expoente de Hurst ( $h$ ).

Na figura 5 observam-se os resultados obtidos por meio da análise DFA para as três séries estudadas. Por meio dessa técnica, calcula-se a função de flutuação  $F'(s)$  para diferentes escalas temporais  $s$  e; quando a série apresenta correlações de longo alcance,  $F'(s)$  apresenta um comportamento do tipo lei de potência ( $F'(s) = s^h$ , com  $h \neq 0,5$ ). Neste contexto, diz-se que a série apresenta fractalidade ou auto-similaridade.



**Figura 05** – Resultados das análises DFA nas séries das intensidades sonoras para uma série de ruídos e duas séries musicais, uma de gênero rock e outra de gênero clássico. A linha contínua, em vermelho, é um ajuste linear dos dados, cujo coeficiente angular é numericamente igual ao valor do expoente de Hurst ( $h$ ). A escala  $s$  é medida em segundos.



A linha reta, em vermelho, corresponde a um ajuste linear, cujo coeficiente angular  $h$  é numericamente igual ao expoente de Hurst. Além disso, nota-se que o valor de  $h$  para a série de ruído branco é  $0,5$ , indicando que os dados da série não tem correlações. Por outro lado, observa-se que para as séries do gênero rock e do gênero clássico os valores de  $h$  são maiores que  $0,5$ . Esse fato indica que as duas séries apresentam correlações de longo alcance e, conseqüentemente, uma memória de longa duração. Porém, quando se compara os valores de  $h$  entre as duas séries musicais, nota-se que a série do gênero clássico tem o seu valor de  $h$  mais próximo de  $1$  ( $h = 0,95$ ), enquanto a série do gênero rock, o valor do expoente  $h$  é  $0,78$ . Diante desses valores, pode-se afirmar que os dados mais correlacionados são aqueles da série do gênero clássico. Além disso, assume-se que a série pertence a uma música mais elaborada (RO, 2009).

## CONSIDERAÇÕES FINAIS

Neste trabalho, apresentamos a aplicação do método *Detrended Fluctuation Analysis* – DFA na identificação de fractalidades e correlações em séries temporais. Descrevemos as principais propriedades de uma série temporal. E, por meio do expoente de Hurst, obtido da análise DFA, podemos identificar fractalidades, correlações de longo alcance e, conseqüentemente, a presença de memória de longa duração numa série. Realizamos uma descrição do algoritmo DFA de uma forma que seja possível sua implementação computacional.

Além disso, o método DFA foi aplicado em três séries temporais. A primeira delas era composta por dados não correlacionados (ruído branco), e o expoente de Hurst ( $h$ ) encontrado foi  $0,5$ . Esse valor para  $h$  já era esperado e está de acordo com a literatura. Numa comparação visual, notamos que a série da canção do gênero *rock* assemelha-se muito mais à série dos ruídos que a série da música do gênero clássico. As outras duas séries estudadas eram formadas por sons musicais. Uma do gênero *rock*, extraída da canção Lutar Pelo Que é Meu da banda Charlie Brown Jr. Na série do gênero *rock*, o valor do expoente de Hurst encontrado foi  $h = 0,78$ , enquanto para a série extraída do Concerto N. 1 in E (La Primavera) de Antonio Vivaldi (música do gênero clássico), o expoente de Hurst obtido foi  $h = 0,95$ . Este último expoente indica que a série obtida da música de Antonio Vivaldi é mais correlacionada que a canção de expoente  $h = 0,78$ .

Do ponto de vista de originalidade, a análise DFA, atualmente, vem sendo cogitado por diversos autores como um possível método para determinar o quão elaborado é um certo som musical.

Como perspectivas de trabalhos futuros, é muito interessante estudar as distribuições de probabilidades das séries das amplitudes sonoras, e usar algum parâmetro, para comparar essas distribuições com os respectivos expoentes de Hurst das séries das intensidades, e assim, encontrar um meio quantitativo de classificar músicas dos mais diversos gêneros.





## REFERÊNCIAS

- ADDISON, Paul S. *Fractals and Chaos: An Illustrated Course*. 1. Ed. London, CRC Press, 1997.
- BULDYREV, S. V.; GOLDBERGER, A. L.; HAVLIN, S.; MANTEGNA, R. N.; MATSA, M. E.; PENG, C. K.; SIMONS, M. and STANLEY, H. E. *Long-range correlation-properties of coding and noncoding DNA-sequences - Genbank analysis*, Phys Rev E, v 51, p. 5084, 1995.
- HALLIDAY, David; WALKER, Jearl. **Fundamentos de Física: Gravitação, Ondas e Termodinâmica**. Volume 2. 9. Ed. Rio de Janeiro, 2012.
- Hu, K.; Ivanov, P.-Ch.; Chen, Z.; Carpena, P. and Stanley, H. E. *Effect of trends on detrended fluctuation analysis*, Phys. Rev. E, v 64, 2001.
- KANTELHARDT, J. *Detecting long-range correlations with detrended fluctuation analysis*, Physica A, v 95, p. 441, 2001.
- KANTELHARDT, J. W.; BERKOVITS, R. S.; HAVLIN, and BUNDE, A. *Multifractal detrended fluctuation analysis of nonstationary time series*. Physica A, Amsterdam, v. 316, p. 87-114, 2002.
- MANDELBROT, B. *The Fractal Geometry of Nature*. New York: W.H. Freeman and Company. 1983.
- MORETTIN, Pedro A.; Toloi, Clélia M. C. **Análise de Series Temporais**. 2. Ed. São Paulo: Blucher, 2006.
- Peng, C. K.; Buldyrev, S. V.; Havlin, S.; Simons, M.; Stanley, H.E.; Goldberger, A.L. *Mosaic organization of DNA nucleotides*. Phys. Rev. E, v.49, p.1685-1689, 1994.
- RO, W.; KWON, Y. *1/f noise analysis of songs in various genre of music*, *Chaos Solitons Fractals*, v 42, p. 2305, 2009.
- SHUKLA, J. **Predictability in the Midst of Chaos: A scientific basis for climate forecasting**. Science, Washington, D.C., 1998.
- SUTTON, R.T.; ALLEN, M.R. *Decadal predictability of North Atlantic sea surface temperature and climate*. Nature, London, v. 388, p. 563-567, 1997.
- TAQQU, Murad S.; TEVEROVSKY, Vadim and WILLINGER, Walter. *Estimators for long-range dependence: an empirical study*. Fractals, London, v. 3, n. 4, p. 785-798, 1995.