RESEARCH   **Open Access**

Low-frequency ambient noise generator with application to automatic speaker classification

Ricardo Santana and Rosângela Coelho[*]

Abstract

A novel low-frequency $1/f$ ambient noise generator using fractional statistics is proposed in this article. The noise samples are obtained by transformation functions performed on pseudo-random uniform sequences. The $1/f$ spectrum representation achieved for the generated noise samples, shows that this proposition is very promising for the investigation of the low-frequency noise effect in signal processing techniques, devices and systems. It is also demonstrated that it can be useful to serve as background ambient noise in speaker classification applications.

Keywords: $1/f$ ambient noise, Fractional Brownian noise, Low-frequency statistics, Speaker classification

Introduction

In the last decades, the presence of low-frequency or $1/f$ noise has been widely observed in such a variety of systems [1,2]. In particular, $1/f$-spectra acoustic noise has been measured in ocean noise [3], music [4] and speech [5]. Noisy environments can severely degrade the performance of speech and speaker classification applications [6-8]. These background noise sources can have different temporal and spectral statistics. Therefore, $1/f$ acoustic noise shall be considered to achieve robust signal processing techniques.

Noises are random processes described by the shape of its power spectral density (PSD). The PSD of noises [1,9] is defined by $S(f) \approx \frac{1}{f^\beta}$ with $0 \leq \beta \leq 2$. Generally, the PSD shape family is achieved by filtering Gaussian white noise (fgwn) sequences using digital finite impulse response (DFIR) filters and signal processing techniques [10-12]. However, the wide-sense stationary can only be measured for very long sample sequences. Mandelbrot and Van Ness [9] showed that the $1/f$ noise statistics can be accurately represented by the fractional Brownian motion (fBm). fBm is defined as a non-stationary stochastic process. Nevertheless, the shape of the PSD and the β exponent can be quasi-stationary if the observed time is short compared to the process life time [1,13]. And thus,

it enables the application of the estimation theory for $1/f$ processes [14,15].

$1/f$ fractional noise has $S(f) \propto f^{1-2H}$, where $1/2 < H < 1$ is the Hurst parameter [16]. The H parameter is described by the slow-decaying rate of the auto-correlation function (ACF) of the noise samples. It represents the low-frequency or scaling invariance degree of the fractional noises and it is frequently close to 1.

This article proposes the generation of $1/f$ ambient noise samples based on the fBm statistics. In the present approach, the $1/f$ spectral behavior is obtained from the ACFs of the noise samples generated by the fBm process. The $1/f$ ambient noise sample generation is based on transformation functions performed on uniform random sequences. These functions are defined by the successive random addition algorithm using the midpoint displacement (SRMD) technique [17]. In a previous study, these transformation functions were successfully evaluated for a low-frequency optical noise samples generation [18].

The solution presented for the SRMD algorithm to generate the $1/f$ acoustic noise samples, is also implemented in a high-speed field-programmable gate array (FPGA) Development Kit. Each noise output value is then pulse coded modulation (PCM) encoded/quantized and sampled at 8 KHz to produce the ambient noise levels. For the experiments, it is considered the real or natural $1/f$ Airport [19] and Airplane [20] ambient noises and also an artificial Pink [20] noise. The validation results

*Correspondence: coelho@ime.eb.br.
Electrical Engineering Department, Acoustic Signal Processing Laboratory of the Military Institute of Engineering (IME), Rio de Janeiro, RJ 22290-270, Brazil

include the estimation of the main parameters or statistics (β exponent, H, mean (μ), variance (σ^2) and Kurtosis (K)), the PSD and the heavy-tail distribution (HTD) curves and the Bhattacharyya distance (B_d). These results are obtained from the real and the generated noise samples. For the experiments, $1/f$ sample sequences are also generated by filtering a Gaussian white noise using the Al-Alaoui transfer function [21]. Furthermore, the performance of the proposed $1/f$ acoustic ambient noise generation is evaluated for a speaker identification task considering different signal to noise ratio (SNR) values.

The rest of the article is organized as follows. Section "$1/f$ fractional Brownian noise: an overview" gives an overview of the $1/f$ fractional Brownian noise and describes the SRMD technique. Section "Implementation setup" introduces the implementation setup of the proposed $1/f$ ambient noise generator. The main validation results are reported and discussed in Section "Validation results and discussion". The speaker classification task and the related results are shown in Section "Speaker classification experiments". Finally, Section "Conclusion" presents the main conclusions of this work.

1/f fractional Brownian noise: an overview

For any instant $t > 0$, $X_H(t)$ is a fractional random function with Gaussian independent increments [9]. The fBm is known as the unique Gaussian H-self-similar with self-similarity parameter and stationary increments (sssi) random process. The variance of the independent increments is proportional to its time interval accordingly to

$$\text{Var}[X_H(t_2) - X_H(t_1)] \propto |t_2 - t_1|^{2H} \qquad (1)$$

for all instants t_1 and t_2 and,

1. $X_H(t)$ has stationary increments.
2. $X_H(0) = 0$ and $E[X_H(t)] = 0$ for any instant t.
3. $X_H(t)$ presents continuous sample paths.

In other words, its statistical characteristics hold for any time scale. Thus, for any τ and $r > 0$,

$$[X_H(t + \tau) - X_H(t)] \overset{d}{\approx} r^{-H}[X_H(t + r\tau) - X_H(t)] \quad (2)$$

where $\overset{d}{\approx}$ means similar in distribution and r is the random process scaling factor. Note that $X_H(t)$ is a Gaussian process completely specified by its mean, variance, H parameter. The ACF of $1/f$ X_H, i.e., $1/2 < H < 1$ is

$$\rho_X(k) = \frac{\sigma_X^2}{2}[(k+1)^{2H} - 2|k|^{2H} + |k-1|^{2H}] \qquad (3)$$

for $k \geq 0$ and $\rho_X(k) = \rho_X(-k)$ for $k < 0$. In the present proposition, the spectral density is derived from the ACF of the $1/f$ fBm noise samples defined in (3). This is ensured by the PSD and ACF exponents that are both related to the H parameter.
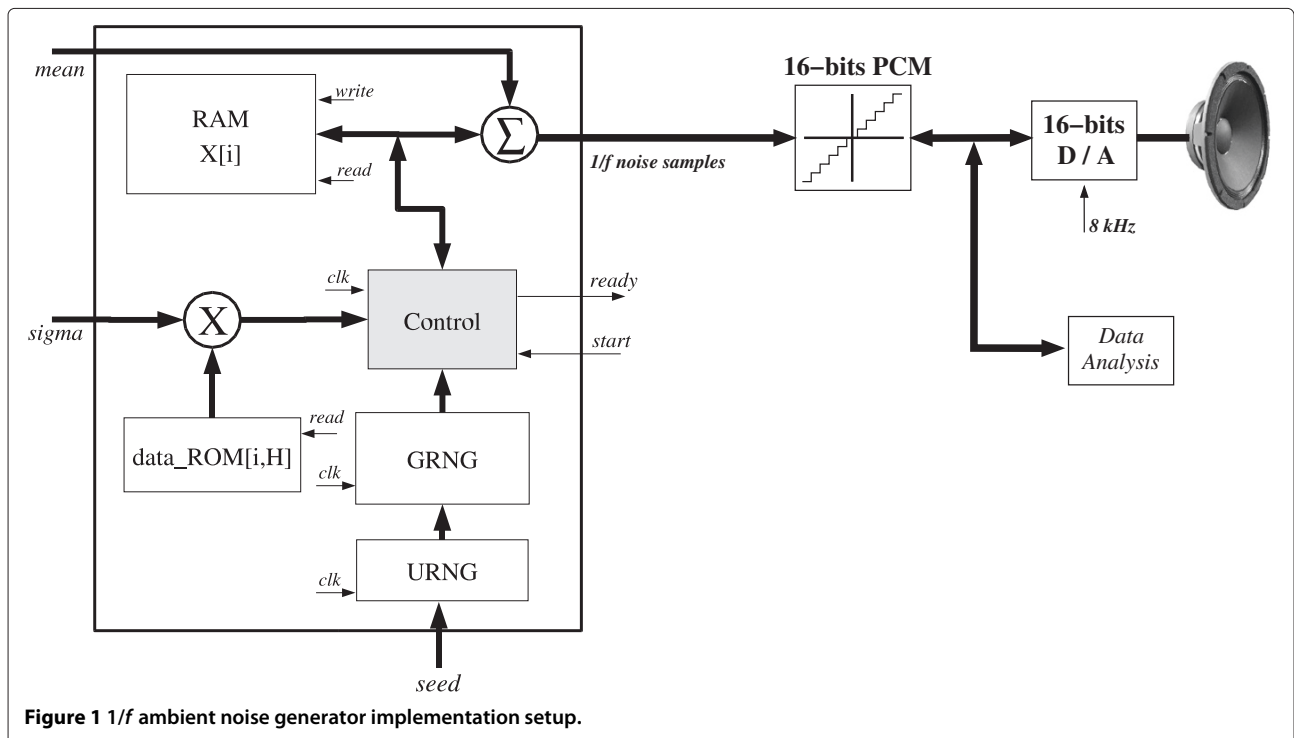

Figure 1 1/f ambient noise generator implementation setup.

SRMD

Considering a time index t defined at the interval $[0, 1]$, the SRMD algorithm establishes that setting $X(0) = 0$ and $X(1)$ as a Gaussian random variable (RV) with zero-mean and variance σ^2 then,

$$\text{Var}[X(1) - X(0)] = \sigma^2. \tag{4}$$

and

$$\text{Var}[X(t_2) - X(t_1)] = |t_2 - t_1|\sigma^2 \tag{5}$$

for $0 \leq t_1 \leq t_2 \leq 1$. To achieve this property a random offset displacement (D_i) with zero-mean and variance $\delta_i^2 = 1/2^{-(i+1)}\sigma^2$, must be added to the noise sample. For example, the $X(1/2)$ value is obtained by the interpolation of $X(0)$ and $X(1)$ with variance $\delta^2/2^{2H+1}$. Several iterations are then proceeded to compose a $1/f$ fBm noise sample sequence. In order to find stationary increments, after the midpoints interpolation, a D_i of a certain variance, $\propto (r^n)^{2H}$ (r is the scaling factor), is applied to all points (time increments) and not just the midpoints. The maximum number of iterations is defined by $N = 2^{\text{maxlevel}}$ where *maxlevel* is generally applied in the interval [0,16] [9]. The other SRMD inputs are the standard-deviation and the H parameter.

Implementation setup

Figure 1 illustrates the experimental setup used for the demonstration of the proposed $1/f$ ambient noise generator. The mean, variance and H values estimated from the real ambient and the artificial Pink noises are used as target parameters. The $1/f$ fBm noise generator is composed by the following main blocks: Uniform random number (URNG); Gaussian random number (GRNG); data_ROM[i, H]; Control and $X[i]$ RAM memory. The generation of the Gaussian RVs is based on transformation functions performed on Uniform variables using the Box-Muller method [22]. The Gaussian random number generator (GRNG) block provides the mean value for the *Gauss(.)* function.

Table 1 β and H estimation results and the β estimation error

Noise	β	$\varepsilon(\beta)$	H
Airplane (real)	1.13	0.124371	0.889
Airplane (proposed)	1.14	0.076390	0.890
Airplane (fgwn)	1.10	0.061164	0.862
Airport (real)	0.89	0.411874	0.882
Airport (proposed)	0.90	0.079132	0.891
Airport (fgwn)	0.81	0.076929	0.813
Pink (real)	1.02	0.041085	0.919
Pink (proposed)	1.01	0.065109	0.915
Pink (fgwn)	0.89	0.088994	0.847

Table 2 Kurtosis, mean and variance statistics results

Noise	K	μ	σ^2
Airplane (real)	2.94	0.001694	0.004665
Airplane (proposed)	3.04	0.001679	0.004665
Airplane (fgwn)	2.98	0.001929	0.004664
Airport (real)	3.11	0.000021	0.003434
Airport (proposed)	3.04	0.000031	0.003381
Airport (fgwn)	2.98	0.000087	0.003428
Pink (artificial)	3.02	0.001402	0.000945
Pink (proposed)	3.03	0.001445	0.000818
Pink (fgwn)	2.99	0.001455	0.000948

The $X[i]$ has $i = 2^{maxlevel}$ increments or noise levels. Uniform random number block was coded to produce 32-bit uniformly distributed samples with periodicity 10^{10}. The linear feedback shift registers (LSFRs) are started by *seed* values to produce the pseudo-random sequences. The data_ROM block performs the computation of the SRMD algorithm. The data_ROM is indexed by i and H and it is defined by data_ROM$[i,H] := \frac{\text{delta}[i]}{\text{sigma}} = (\frac{1}{2})^{iH}\sqrt{\frac{1}{2}}\sqrt{1 - 2^{2H-2}}$. The delta[$i$] values are stored and addressed by the i and H indexes. This would be prohibitive due to the large amount of memory resources needs for storing a wide range of delta[i] values. However, H values can be represented with only two-digits after the decimal point. Thus, 1,000 H values are necessary for each iteration of the SRMD algorithm. Since $0 \leq$ maxlevel $\leq$ 16, delta[i] vector can have a maximum of 16,000 elements. Hence, 1.6% of the ROM memory resource was needed to store the delta[i] vector. In fact, $1/f$ noises have $1/2 < H < 1$ (close to 1) and hence the memory needs can be reduced. This second memory block is used to store the output sample vector ($X[i]$). The binary representation of each $X[i]$ noise output was truncated to 16-bit wide. The main functions of the control block are: Read the GRNG block output (Gaussian samples); Read the data_ROM values according to the selected values indexed by i and H; Evaluate the delta by multiplying previous data_ROM data to standard-deviation (sigma); Fill the initial values of the $X[i]$ vector with the computed sigma $*$ Gauss values; Perform the *loops* iterations (one *while* and two *for*'s); Read fBm output noise sample levels from $X[i]$ vector. The $1/f$ noise SRMD block implementation required ten digital signal processor (*dsp*) blocks and six phase locked

Table 3 B_d results

Noise	Proposed	fgwn
Airplane	0.0169	0.0663
Airport	0.0209	0.0969
Pink	0.0382	0.0372

loop (PLL) used for clock generation to achieve the target noise sample output rate. Following, each noise output value is PCM encoded/quantized and sampled at 8 KHz to produce the ambient noise levels. This sampling rate is necessary for the speaker classification experiments.

Besides the real ambient noises and the artificial Pink noise, samples obtained by filtering a Gaussian white noise are considered for the validation of the proposed method. The applied method uses the Al-Alaoui digital integrator transfer function [21] with $\beta/2$ as the fractional order exponent to compose the transfer function $H(z)$,

$$H(z) = \left[\frac{7T}{8}\frac{(1+z^{-1}/7)}{(1-z^{-1})}\right]^{\beta/2}, \qquad (6)$$

where T is the sampling period. The filter coefficients are obtained by the convolution $h(k) = a(k) * b(k)$, where $a(k)$ and $b(k)$ are the first $N/2$ terms obtained by expanding, respectively, the numerator and denominator of (6) in power series [12].

Validation results and discussion
β and *H* estimation results

The β exponent is estimated from the linear regression applied to the PSD function curves. Table 1 shows the β exponent, the mean square error (MSE) of the β estimation, and the H results obtained from the real and artificial noises, and from the noise samples generated by the proposed and the fgwn methods. The results are presented for 320,000 samples since this is the size of the real ambient noise sequences. For the H estimation it is used the wavelet-based method [23] with 12 Daubechies filters and the 4–12 scale range. It can be seen that the β exponent values of the noise samples obtained from the proposed method, are very close to the values of the real ambient noises.

It can be noted that for the artificial Pink noise samples, the proposed and fgwn methods achieve quite similar H target values, i.e., the low-frequency statistics. However, the H results estimated from the noise samples generated by the proposed method, are much closer to the H values of real ambient noises.
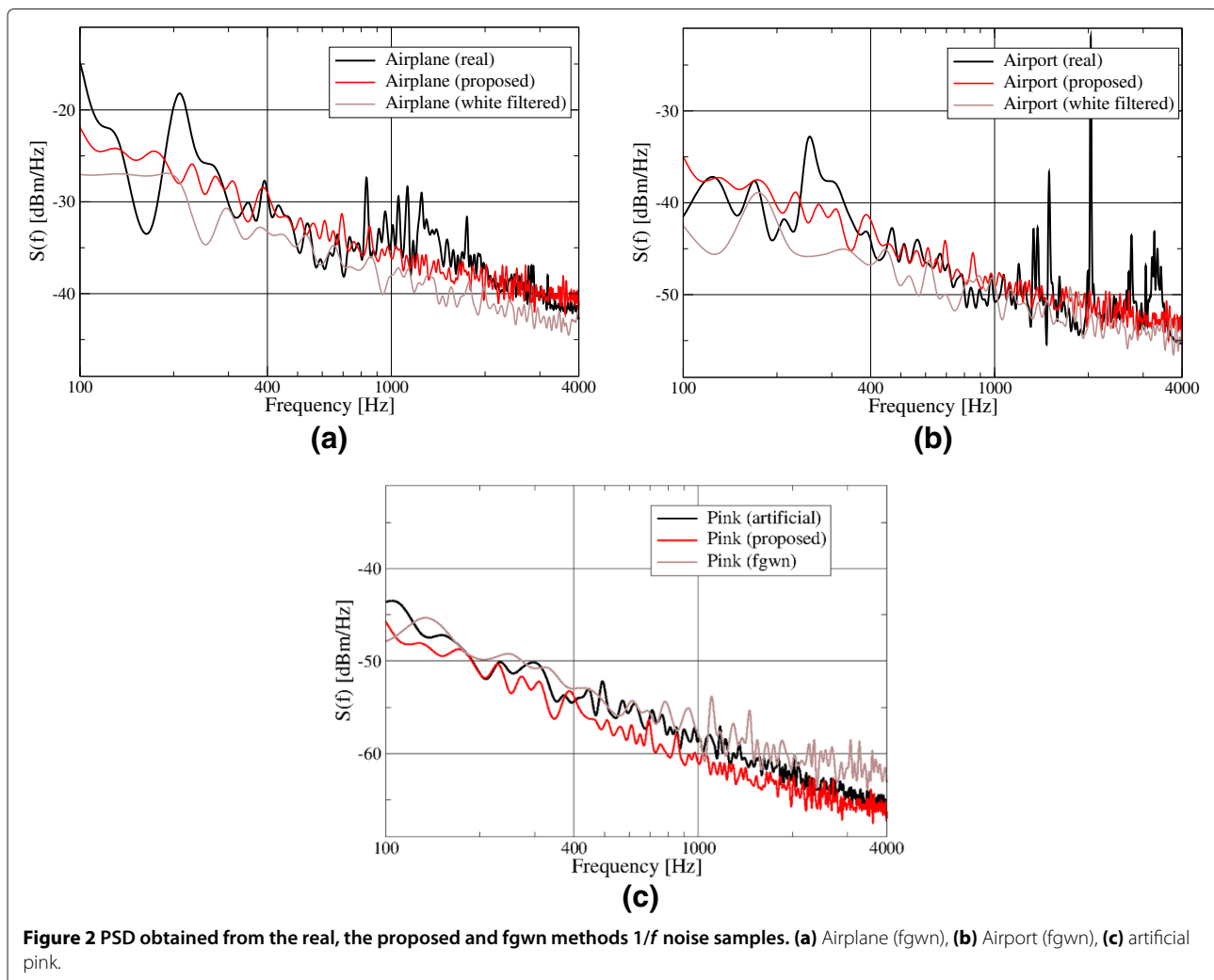

Figure 2 PSD obtained from the real, the proposed and fgwn methods 1/*f* noise samples. **(a)** Airplane (fgwn), **(b)** Airport (fgwn), **(c)** artificial pink.

Kurtosis, mean, variance statistics

Kurtosis measures the skewness of a sample from a Gaussian distribution. The K, mean and variance estimation results of the noise sequences are presented in Table 2. As expected, the K values are close to 3. Thus confirming that the noise samples are Gaussian distributed.

Bhattacharyya distance

The Bhattacharyya distance measures the separability between two sample sequences with Gaussian distribution and is defined by

$$B_d = \frac{1}{2} \ln \frac{\frac{|C_i + C_j|}{2}}{|C_i|^{1/2}|C_j|^{1/2}}$$
$$+ \frac{1}{8}(\mu_i - \mu_j)^T \left(\frac{C_i + C_j}{2} \right)^{-1} (\mu_i - \mu_j) \qquad (7)$$

where μ_i is the mean vector and C_i is the covariance matrix of class $i = 1, 2$. The B_d are measured between the generated sequences and the corresponding Airplane, Airport and Pink noises. It can be seen from Table 3 the Pink

noise samples distribution produced by both methods, is very similar to the distribution of the artificial Pink noise. However, the samples distribution obtained from the proposed method are much similar to the distribution of the real ambient noises.

PSD results

The power spectral densities obtained from the real and the generated $1/f$ acoustic noise samples are presented in Figure 2. The PSDs were measured using a high-performance 300 MHz bandwidth spectrum analyzer. These results demonstrate the slow-decaying (3 dB/octave) behavior of the PSD shape of the $1/f$ noises. It can also be seen that the proposed method better represents the PSD behavior of the real acoustic noises.

HTD results

Figure 3 illustrates the HTD ($P[X > x]$) curves of the Airplane, Airport and Pink noise samples achieved by the proposed and fgwn methods. The HTD curves obtained from the real and the proposed method demonstrate that
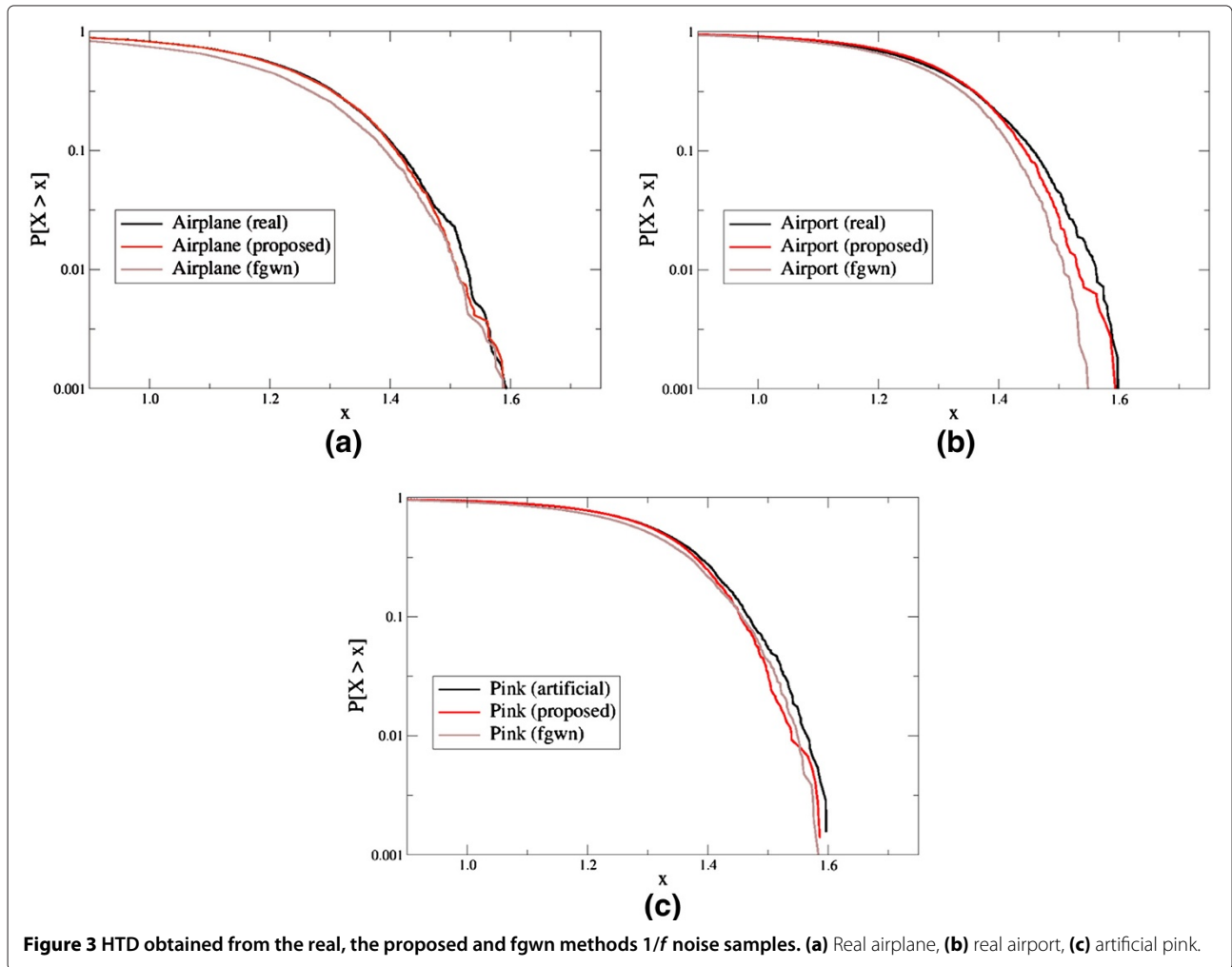

Figure 3 HTD obtained from the real, the proposed and fgwn methods 1/f noise samples. (a) Real airplane, **(b)** real airport, **(c)** artificial pink.

they exhibit very close tails. This also confirms the H results (see Table 1) obtained from the proposed solution.

Speaker classification experiments

In a speaker identification process, a speech utterance has to be identified as to which of the registered speakers it belongs. For the experiments, the speech utterances were corrupted with the real and generated noise samples. For the speaker identification were considered the mel-cepstral coefficients (MFCC) and the Gaussian mixed model (GMM) [24] which are respectively, the most commonly used speech features and classifier employed in speaker recognition tasks.

A mixture of Gaussian probability densities is a weighted sum of M densities, and is given by

$$p(\vec{x}|\lambda) = \sum_{i=1}^{M} p_i b_i(\vec{x}) \tag{8}$$

where $\vec{x}$ is a random vector of dimension D, $b_i(\vec{x})$, $i = 1,\ldots,M$, are the Gaussian density components, and p_i, $i = 1,\ldots,M$, are the mixture weights.

Each component density is a D variate Gaussian function of the form

$$b_i(\vec{x}) = \frac{e^{\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T K_i^{-1}(\vec{x}-\vec{\mu})\right)}}{(2\pi)^{\frac{D}{2}}\sqrt{|K_i|}} \tag{9}$$

with mean vector $\vec{\mu}_i$ and covariance matrix K_i, where T denotes the transpose operation and $|.|$ is the determinant. The GMM (λ) is parametrized by the mean vectors, covariance matrices, and mixture weights. The model parameters are estimated for a set of training data as the ones that maximize the likelihood of the GMM. The expectation-maximization (EM) algorithm [24] is used for the model parameters estimates. Considering a sequence of T independent training vectors $X = \{\vec{x}_1,\ldots,\vec{x}_T\}$, the normalized log-likelihood of the GMM is

$$\log p(X|\lambda) = \frac{1}{T}\sum_{t=1}^{T} \log p(\vec{x}_t|\lambda). \tag{10}$$

The decision rule of the speaker identification system chooses the speaker model for which this value is maximum.

Speaker identification accuracy results

The speaker identification task evaluation is performed on the KING speech corpus. This is composed of conversational sessions of speech recorded by 49 male speakers. For the experiments, five sessions are used resulting in 100 s average of speech per speaker, after silence removal. Three of these sessions (60s) are applied for the speaker model training. The remaining two sessions (40s), are used to evaluate the identification accuracies.

The speaker classification results are presented for test duration of 5 and 1 s. The real and generated $1/f$ noise samples are added to the speech utterances serving as background ambient noise. For the investigation, it is also considered the SNR 0 dB, 5 dB, 10 dB, 15 dB, and 20 dB to evaluate the system under different noisy conditions. For the identification task it was considered speech feature vectors with 25 MFCCs, extracted from 20 ms speech frames, and $M = 32$ GMM components. The speaker identification accuracies are shown in Figure 4. The results show that the generated $1/f$ noise produced similar effect when compared to the real ambient noise. This means that it could be applied as artificial background noise.

Conclusion

A new low-frequency $1/f$ ambient noise generator using fractional statistics is described in this article. The PSD shape of the $1/f$ generated noise samples is achieved from the ACFs of the noise samples generated with the fBm process. The implementation of the $1/f$ ambient noise
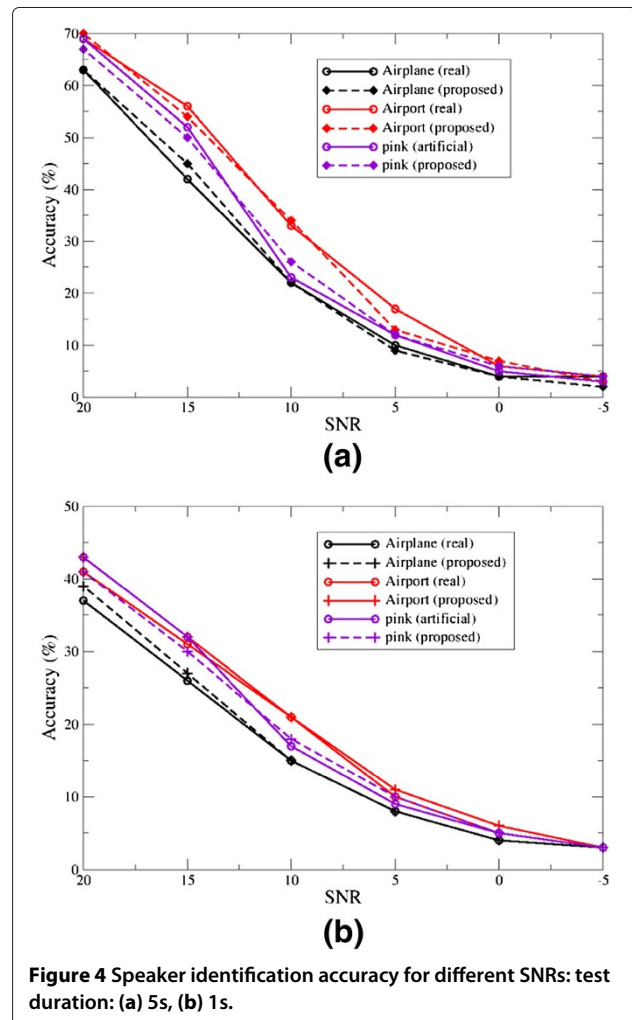

Figure 4 Speaker identification accuracy for different SNRs: test duration: (a) 5s, (b) 1s.

generator enables the validation of the pattern and the PSD representation. It is shown that this proposition is very promising for the investigation of this noise effect in the signal processing techniques. Furthermore, the speaker identification experiments demonstrate that the generated ambient noise samples can be useful to serve as background or additive noise.

Competing interests

The authors declare that they have no competing interests.

References

1. M Keshner, 1/f noise. Proc. IEEE. **70**, 212–218 (1982)
2. F Hooge, 1/f noise. Physica B+. **C83**, 14–23 (1976)
3. A Derzjavin, A Semenov, Ocean ambient low frequency acoustic noise structure in shallow and deep water regions. Journal de Physique IV. **4**, 1269–1272 (1994)
4. R Voss, J Clarke, 1/f noise in music: Music from 1/f noise. J. Acoust. Soc. Am. **63**(1), 258–263 (1978)
5. R Voss, J Clarke, 1/f noise in music and speech. Nature. **258**, 317–318 (1975)
6. Y Gong, Speech recognition in noisy environments: a survey. Speech Commun. **16**, 261–291 (1995)
7. J Ming, T Hazen, J Glass, D Reynolds, Robust speaker recognition in noisy conditions. IEEE Trans. Audio Speech Lang. Process. **15**, 1711–1723 (2007)
8. L Zão, R Coelho, Colored noise based multicondition training technique for robust speaker identification. IEEE Signal Process. Lett. **18**, 675–678 (2011)
9. B Mandelbrot, J Van Ness, Fractional brownian motions, fractional noises and applications. SIAM Rev. **10**, 422–437 (1968)
10. M Deriche, A Tewfik, Signal modeling with filtered discrete fractional noise processes. IEEE Trans. Signal Process. **41**(9), 2839–2849 (1993)
11. C Tseng, S Pei, S Hsia, Computation of fractional derivatives using fourier transform and digital fir differentiator. Signal Process. **80**(1), 151–159 (2000)
12. Y Ferdi, A Taleb-Ahmed, M Lakehal, Efficient generation of $1/f^\beta$ noise using signal modeling techniques. IEEE Trans. Circ. Syst. **55**, 1704–1710 (2008)
13. F Hooge, Discussions of recent experiments on 1/f noise. Physics. **60**, 130–144 (1972)
14. S Yousefi, J Jaldén, T Eriksson, Linear prediction of discrete-time 1/f processes. Signal Process. Lett. **17**(11), 901–904 (2010)
15. B Ninness, Estimation of 1/f noise. IEEE Trans. Inf. Theory. **44**, 32–46 (1998)
16. E Hurst, Methods of using long–term storage in reservoirs. Proc. Inst Civil Eng, 519–543 (1956)
17. M Barnsley, R Devaney, B Mandelbrot, H Peitgen, D Saupe, R Voss, *The Science of Fractal Images*. (New York, USA: Springer-Verlag, 1988)
18. L Zão, R Coelho, Low-frequency optical noise generator using fractional statistics. Electron Lett. **46**, 1072–1074 (2010)
19. FreeSFX, Airport ext busy tarmac, Ambiences/Background Sound Effects. [http://www.freesfx.co.uk/soundeffects/airports/]; 2009
20. A Varga, H Steeneken, M Tomlinson, M Jones, The noisex-92 study on the effect of additive noise on automatic speech recognition. Technical Report of Defence Evaluation and Research Agency [http://spib.rice.edu/spib/1992]
21. M Al-Alaoui, Novel digital integrator and differentiator. Electron. Lett. **29**, 376–378 (1993)
22. G Box, M Muller, A note on the generation of random normal deviates. Ann. Math. Stat. **29**, 610–611 (1958)
23. P Flandrin, Wavelet analysis and synthesis of fractional brownian motion. IEEE Trans. Inf. Theory. **38**, 910–917 (1992)
24. D Reynolds, R Rose, Robust text-independent speaker identification using gaussian mixture speaker models. IEEE Trans. Speech Audio Process. **3**, 72–83 (1995)