

A new spectrum feature discovered for a category of signals produced by massive and random micro-sources

Xiaodong Zhuang¹, and Nikos Mastorakis^{2,*}

¹Electronics & Information College, Qingdao University, Qingdao 266071, China

²Department of Electrical Engineering and Computer Science, Hellenic Naval Academy, Piraeus, Greece

Abstract. A statistical study is implemented on the short-time spectrum of one main category of random signals. For the signals with massive and random micro-sources, a new statistic feature of the short-time amplitude spectrum is discovered, which reveals the relationship between the amplitude's average and its standard for each frequency component. Moreover, the association between the amplitude distributions for different frequency components is also studied. A model representing such association is presented, which accords well with the statistic feature discovered. The analysis result has potential application in signal classification, and also in the study of system characteristics underlying the observed signal.

1 Introduction

From the viewpoint of computing, a signal processing method is a computing process taking the signal as input. Due to the large variety of signal types, it is important to identify the signal characteristics (or category) for the selection of a proper method for the specific task [1, 2]. Signals are the observation result of a system. In practice, a system usually consists of massive micro elements. Each element contributes to the generation of the signal. These elements are named "micro-sources" in this paper. The observed signal is the accumulation effect of all the micro-sources. The relationship between these micro-sources determines the basic characteristic of the signal, which can act as the basis for signal classification. In some systems, the micro-sources are strongly associated, such as different parts of a tuning fork moving synchronously in vibration. Other systems have loosely coupled micro-sources and these sources move randomly, such as the air turbulence generating sound, or some friction producing sound [3-7].

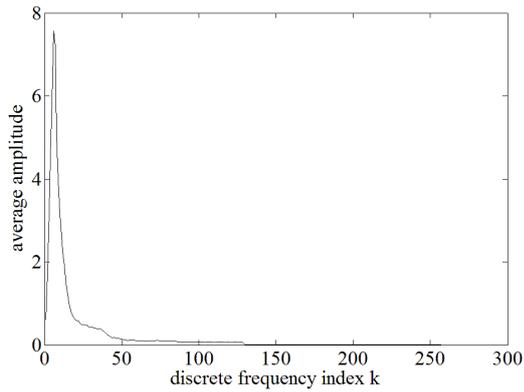
In this study, the statistic characteristic of the spectrum is studied for the later category of signals. For this type of signals, the signal source can be decomposed to massive and random micro-sources. A novel statistical feature for short-time amplitude spectrum is discovered for this category of signals. The feature is named "consistent standard deviation coefficient". Moreover, the relationship between the amplitude probability distributions of two different frequency components is investigated, based on which a model is proposed representing such relationship. The validity of the new model is verified by analysis, with the discovered statistic feature as direct evidence.

2 Statistical study of spectrum for a category of random signals

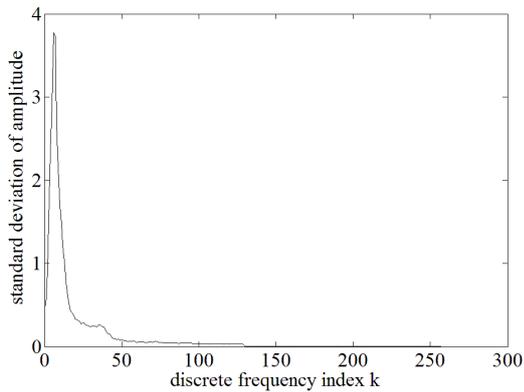
The signals studied are sustained signals captured in real world, with sufficient signal length for statistical study. The signals include the sound of wind, plane engine room, electric fan, rain, river, etc. In the study, each signal is divided into frames, and DFT is performed for each windowed frame. A Hamming window is used on each frame. In another word, the short-time Fourier transform (STFT) is used to gather sufficient spectrum data for the statistic study. The amplitude spectrum is studied. In the discrete spectrum, there is finite number of discrete components, and the statistic study is eventually performed for each frequency component individually.

The signals' sample frequency is 22 kHz. In the STFT on each signal, the frame length is set to 512, which corresponds to a time interval of 23.3 ms for 22 kHz sampling frequency. The study is implemented by programming in Matlab. Let ω_k denotes the k -th frequency component in STFT. Due to the randomness of the signal, the amplitude of ω_k also varies randomly in each frame of the signal. Let $\mu(\omega_k)$ and $\sigma^2(\omega_k)$ represent the estimated average and variance of ω_k 's amplitude respectively. And the estimated standard deviation $\sigma(\omega_k)$ is the square-root of $\sigma^2(\omega_k)$. Mathematically, $\mu(\omega_k)$ and $\sigma(\omega_k)$ are two functions, and their curves can be drawn after μ and σ are estimated for each frequency ω_k . Some results of the curves of $\mu(\omega_k)$ and $\sigma(\omega_k)$ are shown in Fig. 1 to Fig. 6. It can be observed evidently that there is clear similarity between the curves of $\mu(\omega_k)$ and $\sigma(\omega_k)$. Such similarity also exists in all the other results of signals studied (signals of massive and random micro-sources), which inspires the investigation of the relationship between $\mu(\omega_k)$ and $\sigma(\omega_k)$ as following.

* Corresponding author: mastor@hna.gr

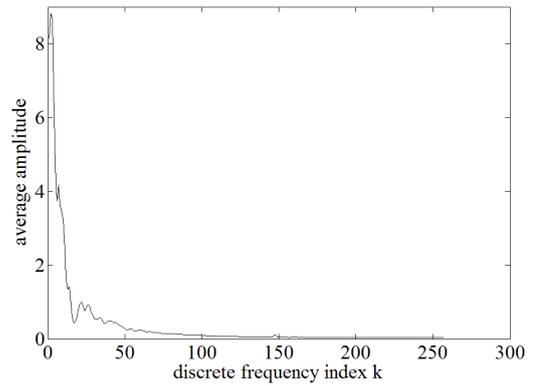


(a) the average amplitude for each ω_k

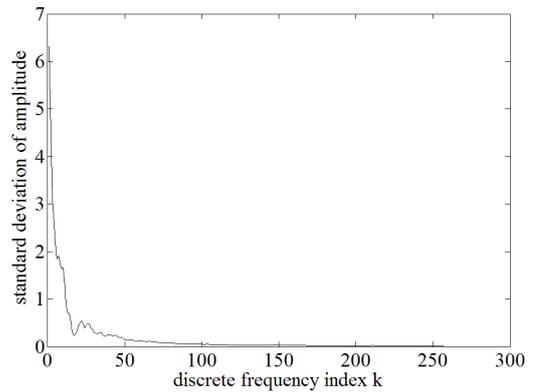


(b) the standard deviation of the amplitude for each ω_k

Fig. 1. The result of $\mu(\omega_k)$ and $\sigma(\omega_k)$ for the signal of wind sound

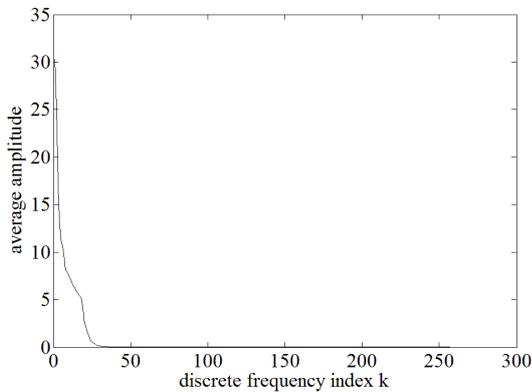


(a) the average amplitude for each ω_k

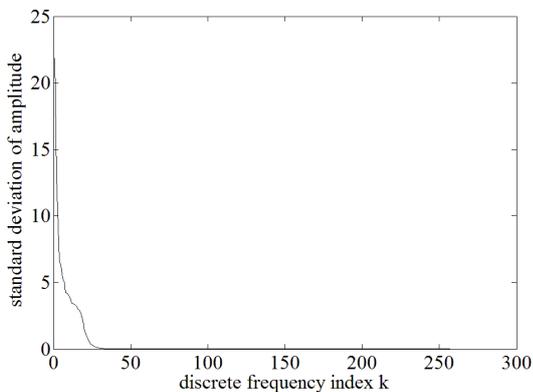


(b) the standard deviation of the amplitude for each ω_k

Fig. 3. The result of $\mu(\omega_k)$ and $\sigma(\omega_k)$ for the signal of electric fan

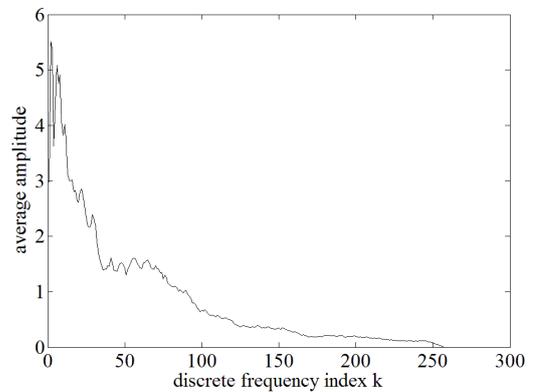


(a) the average amplitude for each ω_k

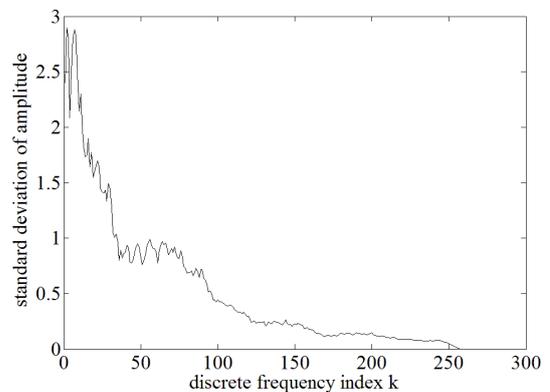


(b) the standard deviation of the amplitude for each ω_k

Fig. 2. The result of $\mu(\omega_k)$ and $\sigma(\omega_k)$ for the signal of engine room



(a) the average amplitude for each ω_k



(b) the standard deviation of the amplitude for each ω_k

Fig. 4. The result of $\mu(\omega_k)$ and $\sigma(\omega_k)$ for the signal of the rain

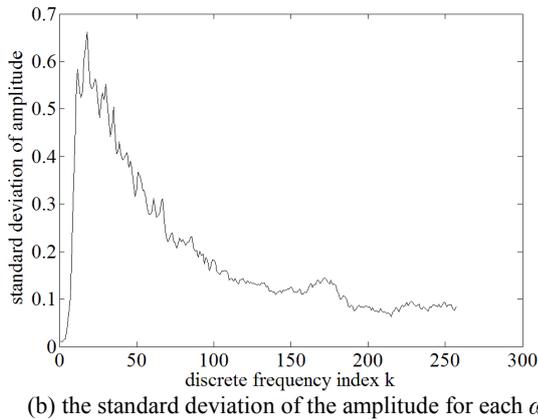
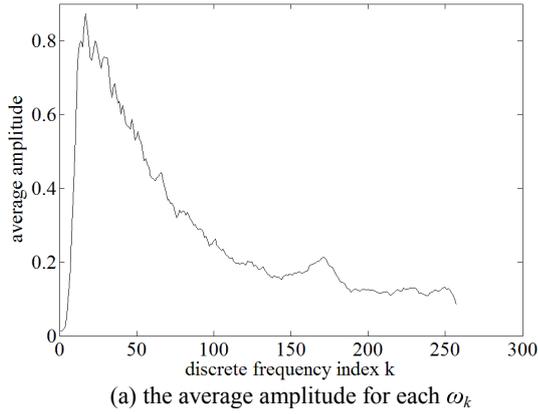


Fig. 5. The result of $\mu(\omega_k)$ and $\sigma(\omega_k)$ for the signal of a river

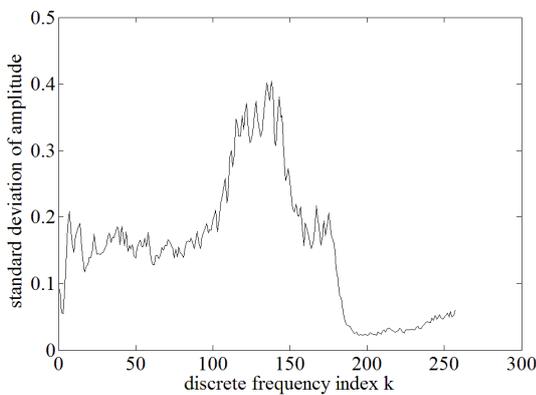
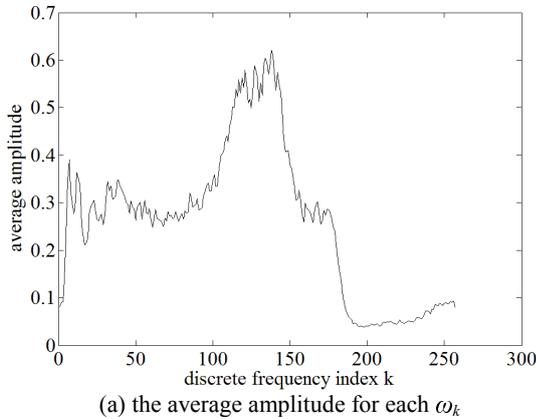


Fig. 6. The result of $\mu(\omega_k)$ and $\sigma(\omega_k)$ for the signal of scratching a paper

Besides the above experimental results, the relationship between $\mu(\omega_k)$ and $\sigma(\omega_k)$ is quantitatively verified by calculating the correlation coefficient between the two curves of $\mu(\omega_k)$ and $\sigma(\omega_k)$. The correlation coefficient is calculated in a discrete form:

$$\rho_{\sigma\mu} = \frac{\sum_{k=1}^N \sigma(\omega_k) \cdot \mu(\omega_k)}{\sqrt{\sum_{k=1}^N \sigma^2(\omega_k)} \cdot \sqrt{\sum_{k=1}^N \mu^2(\omega_k)}} \quad (1)$$

where N is the number of discrete frequencies in the discrete spectrum. Some of the experimental results are shown in Table 1. The correlation coefficients between $\mu(\omega_k)$ and $\sigma(\omega_k)$ are calculated for different signals. The correlation coefficients between $\mu(\omega_k)$ and $\sigma(\omega_k)$ are much close to 1.0. Consider the unavoidable error caused by the instability of the signals, and also the noise introduced in the signal capture process, the results indicate that $\mu(\omega_k)$ and $\sigma(\omega_k)$ are strongly related by a linear proportional relationship, which is a new statistic feature discovered for this category of signals.

Table 1. The correlation coefficient ρ between $\mu(\omega_k)$ and $\sigma(\omega_k)$ for the signals studied.

Signal	ρ between $\mu(\omega_k)$ and $\sigma(\omega_k)$
Sound of wind blowing	0.9990
Sound in a plane's engine room	0.9995
Sound of an electric fan	0.9900
Sound of the rain	0.9978
Sound of a river	0.9974
Sound of scratching a paper	0.9957

Because the parameter of “standard deviation coefficient” represents the σ to μ ratio, the above statistic feature is named as the feature of “consistent standard deviation coefficient”. In another word, for the signals studied here, the proportional coefficient between the standard deviation and the expectation is consistent for all the frequency components in the short-time amplitude spectrum. This feature can also be expressed by:

$$\sigma(\omega_k) = c_s \cdot \mu(\omega_k) \quad (2)$$

where c_s is the consistent standard deviation coefficient of amplitude for all frequency components. The subscript s means that the value of c_s is for one specific signal. If the signal is changed to another signal, the value c_s may also change. Because the expectation and the standard deviation are two basic statistic of a random variable, the feature of “consistent standard deviation coefficient” indicates that there is certain association between the amplitude probability distributions of different frequency components, which is studied in the next section.

3 The statistical relationship between different frequency components

Based on the spectrum data obtained by STFT, the histogram of amplitude for each frequency component ω_k is computed. The histogram reflects the distribution of random amplitude data for each ω_k , which is closely related to the amplitude probability distribution. Therefore, the amplitude histogram of each ω_k is compared to those of other frequencies, in order to study the relationship between them. In order to study the amplitude distribution of different ω_k without the influence of different average value, the normalized histogram is also computed for each ω_k . First, the average of amplitude for ω_k is computed. After that, each amplitude data of ω_k is divided by that average value as a preprocessing step. The normalized histogram is then computed based on the data after that preprocessing. In the results, the normalized histogram curves obviously converge to one central curve, which indicate the strong association between the amplitude distributions of different ω_k .

Based on such results, a model of amplitude distribution in frequency domain is proposed for the signals studied here. For each signal, the amplitude distributions for different ω_k are supposed of the same type, but with different expectation (or average) values. In another word, there is a prototype distribution function $p_0(a_0)$, from which the amplitude distribution of any ω_k can be derived by varying the expectation.

The above model can also be described mathematically as follows. As a random variable a , the amplitude of some ω_k is modeled as the scaling of a prototype random variable a_0 , whose expectation is 1:

$$a = k \cdot a_0 \quad (3)$$

where k is the scaling parameter. Equation (3) is a mathematical description of the model proposed. In the model, a_0 is the same for each frequency component, but the scaling parameter k may be different for different ω_k .

The proposed model can also find proof from the new discovered statistic feature in Section 2. In the following, the feature of “consistent standard deviation coefficient” can be theoretically induced from the proposed model; in another word, this model accords well with the feature of “consistent standard deviation coefficient” discovered in the experiments. First, consider the probability distribution of a in Equation (3), given $p_0(a_0)$ is the probability distribution of a_0 . According to Equation (3), the expectation of a is :

$$\mu_a = E[a] = E[k \cdot a_0] = k \cdot E[a_0] = k \cdot \mu_0 \quad (4)$$

where μ_0 is the expectation of a_0 . Based on the pdf (probability distribution function) of a variable's function in probability theory, the probability distribution of a can be deduced as:

$$p(a) = \frac{1}{k} \cdot p_0\left(\frac{a}{k}\right) \quad (5)$$

Consider the standard deviation coefficient of a :

$$\frac{\sigma_a}{\mu_a} = \frac{\sqrt{Var(a)}}{\mu_a} = \frac{\sqrt{\int_{-\infty}^{+\infty} (a - \mu_a)^2 p(a) da}}{\mu_a} \quad (6)$$

Considering Equation (4) and (5), Equation (6) can be rewritten as:

$$\frac{\sigma_a}{\mu_a} = \frac{\sqrt{\int_{-\infty}^{+\infty} (a - k \cdot \mu_0)^2 \cdot \frac{1}{k} p_0\left(\frac{a}{k}\right) da}}{k \cdot \mu_0} \quad (7)$$

Then do the variable substitution $a = k a_0$ to the integral on the right side of Equation (7):

$$\begin{aligned} \frac{\sigma_a}{\mu_a} &= \frac{\sqrt{\int_{-\infty}^{+\infty} (k a_0 - k \mu_0)^2 \cdot \frac{1}{k} p_0(a_0) d(k a_0)}}{k \mu_0} \\ &= \frac{\sqrt{k^2 \cdot \int_{-\infty}^{+\infty} (a_0 - \mu_0)^2 \cdot p_0(a_0) da_0}}{k \mu_0} \end{aligned} \quad (8)$$

Remember that the variables a and a_0 represent the amplitude value, which is non-negative. Therefore, k is also non-negative. Then Equation (8) can be rewritten as:

$$\frac{\sigma_a}{\mu_a} = \frac{\sqrt{\int_{-\infty}^{+\infty} (a_0 - \mu_0)^2 \cdot p_0(a_0) da_0}}{\mu_0} \quad (9)$$

Notice that the numerator of the right side of Equation (9) is just the standard deviation of a_0 . Therefore,

$$\frac{\sigma_a}{\mu_a} = \frac{\sigma_0}{\mu_0} \quad (10)$$

Notice that the right side of Equation (10) is constant given the prototype distribution $p_0(a_0)$. Therefore, the standard deviation coefficient of a is consistent whatever the scaling factor k is, which is equal to that of the prototype variable a_0 . This just accords well with the experimental results shown in Section 2. Therefore, the feature of “consistent standard deviation coefficient” supports the model proposed here.

4 Conclusion

The study in the paper indicates the possibility of signal classification by the statistic characteristics in frequency domain. The study is focused on the short-time amplitude spectrum, which is obtained by implementing STFT for the signals of massive and random micro-sources. A new feature of “consistent standard deviation coefficient” is revealed for this category of signals. This feature indicates strong association between amplitude distributions of different frequency components. A new model is proposed to representing such association. In this model, the random variables representing amplitude of every frequency component belong to the same pdf type, but they have different expectations. Moreover, by mathematical analysis, this model accords well with the feature of “consistent standard deviation coefficient”. These results may facilitate signal classification and signal detection, which will be studied in further research. In future work, a specific pdf type will be studied to suit the short-time amplitude spectrum data for

this type of random signals. And other types of signals like quasi-periodic signals in practice will also be studied.

References

1. K. Todros, A. O. Hero, IEEE Transactions on Signal Processing, **63**(5), 1156-1170 (2015)
2. Y. Li, K. M. Wong, IEEE Journal of Selected Topics in Signal Processing, **7**(4), 655-669 (2013)
3. H. Song, Y. He, X. Yan, D. Shang, 2016 IEEE/OES China Ocean Acoustics (COA), 1-8 (2016)
4. Z. Hu, W. Lu, 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 885-889 (2017)
5. B. Zhang, X. L. Chen, D. S. Huang, Z. Lin, 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and Robotics, Automation and Mechatronics (RAM), 345-349 (2017)
6. D. J. Sinder, M. H. Krane, J. L. Flanagan, Proceedings of 16th International Congress Acoustics, **1**, 249-250 (1998)
7. R. Mittal, B. D. Erath, M. W. Plesniak, Annual Review of Fluid Mechanics, **45**, 437-467 (2013)