IEICE Proceeding Series

Information Aggregation via Stochastic Resonance in Moments

Akihisa Ichiki, Yukihiro Tadokoro

Vol. 1 pp. 13-16
Publication Date: 2014/03/17
Online ISSN: 2188-5079

Information Aggregation via Stochastic Resonance in Moments

Akihisa Ichiki and Yukihiro Tadokoro

Toyota Central R&D Labs., Inc.
Yokomichi, Nagakute, Aichi, 480-1192 Japan
Email: aichiki@mosk.tytlabs.co.jp, tadokoro@mosk.tytlabs.co.jp

Abstract—A **stochastic resonance** (SR) for the purpose of signal detection is investigated. SR conventionally refers to the amplification of a weak signal in the average of the output. This means that only the first moment of the output is considered in the conventional SR. In contrast, higher moments of the output are also modulated by the input signal, and include statistically independent information about the signal. We find that such higher moments can exhibit SR behaviors. The **signal-to-noise ratio** (SNR) improves compared with the conventional SR, by the appropriate aggregation of the information obtained through the SR in moments. The aggregation is realized by the method known as **principal component analysis** (PCA). The SNR obtained by PCA also exhibits SR behaviors. We investigate the SR behaviors of moments in a K-valued non-dynamical system.

1. Introduction

Noise has been focused on its negative side in the field of signal processing: a deterioration of the signal detection performance. However, the discovery of the phenomenon nowadays called **stochastic resonance** (SR) has opened the positive exploitation of noise. Benzi *et al.* reported that a weak signal in noise environment can be enhanced by tuning the noise intensity [1, 2]. Since the discovery of SR, the positive application of noise has been widely explored [3, 4]. Signal detection and amplification using SR have been investigated intensively in a variety of nonlinear systems such as the Schmitt trigger [5], the neuron model [6], and SQUID [7].

In this paper, we focus on the signal detection performance that is achieved by the use of SR in moments. An input signal modulates all moments of the output from a nonlinear device. Then, each of the statistically independent moments has independent information about the input signal. We will show that such moments can exhibit SR behaviors. In contrast, the conventional SR refers only to the signal amplification in the first moment of the output. Thus, the SR displayed in this paper is a novel type of SR. Since the first moment is the only statistically independent moment in a 2-valued system, the conventional SR achieves the true signal detection performance in a 2-valued system [6, 8, 9]. On the other hand, in K-valued systems, the potentially achievable high signal detection performance characterized by **signal-to-noise ratio** (SNR) is obtained only when all of the statistically independent moments are concerned. However, even for K-valued systems, the SR behavior of only the first moment is focused in conventional studies [10, 11].

The goal of the present paper is to show that signal detection performance is improved by aggregating the information obtained from the SR in higher moments. Such an aggregation is provided by the method called **principal component analysis** (PCA). In the proceeding section, SR phenomena in moments is presented. Then, the information aggregation using PCA is presented in section 3. The last section is devoted to the concluding remarks.

2. SR in Moments

Throughout this paper, we deal with a K-valued discrete time non-dynamical system [10, 11]:

$$x_n = F(s_n + z_n), \tag{1}$$

where s_n, z_n, and x_n represent an input signal, noise, and the output at time n, respectively. The noise z_n is a white noise: $\langle z_n z_m \rangle \propto \delta_{nm}$, where $\langle \cdot \rangle$ represents an expectation value. Furthermore, the input signal s_n is assumed to be small compared to the noise intensity, since of our interest is the detection of a weak signal. The function F is a K-valued discrete function: $F(X) = a_k$ for $X \in I_k$, where the domains $\{I_k\}$ satisfy $I_k \cap I_l = \phi$ for $k \neq l$ and $\cup_{k=0}^{K-1} I_k = \mathbb{R}$ ($k = 0, 1, \cdots, K-1$). The real constants $\{a_k\}$ are provided as $a_k \neq a_l$ for $k \neq l$. A schematic system-flow diagram is depicted in Figure 1.

In order to analyze the signal detection performance using the information included in moments systematically, we first consider a small constant input signal $s_n = s$ ($|s| \ll 1$) for all n. Then, the system is regarded as outputting a $(K-1)$-dimensional vector $(M_1, \cdots, M_{K-1})$ that is a sequence of the statistically independent $K-1$ empirical moments $M_k = \sum_{n=0}^{N-1} x_n^k / N$ ($k = 1, \cdots, K-1$), where N is the time length of the observation or the number of the sampled outputs. In order to obtain $K-1$ statistically independent moments from the outputs, we assume $N \geq K-1$. Owing to the stationarity of the stochastic process, the estimated input signal $\hat{s} = \hat{s}(x_0, \cdots, x_{N-1})$, which is a function of the output sequence, is invariant under the permutation of the outputs $(x_0, \cdots, x_{N-1}) \to (x_{n_0}, \cdots, x_{n_{N-1}})$. Note that the output sequence $\{x_n\}$ ($n = 0, \cdots, N-1$) is determined by the $K-1$ empirical moments.

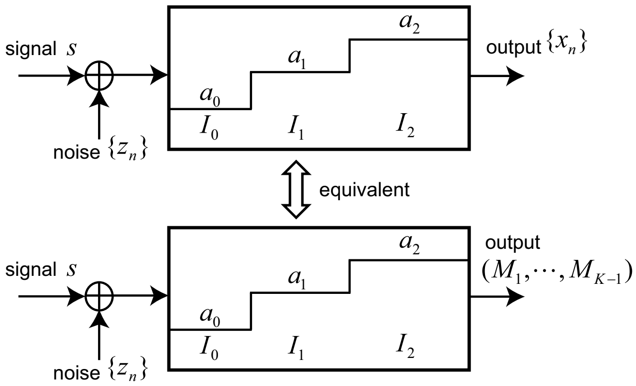

Figure 1: Schematic system-flow diagram of a K-valued non-dynamical system. The system can be regarded as outputting the vector $(M_1, \cdots, M_{K-1})$.

Since the input signal is sufficiently small compared with the noise intensity, the moment M_k can be written as

$$M_k = m_k + \chi_k s + \eta_k + O(s^2). \tag{2}$$

Here, m_k is the expectation value of the k-th order moment without the input s, i.e., $m_k = \langle F^k(z)\rangle$, χ_k is the response coefficient $\chi_k = \partial\langle F^k(s+z)\rangle/\partial s|_{s=0}$, and η_k denotes the fluctuation. In this expression, $\chi_k s$ represents the deterministic signal part of the output in the moment, and η_k corresponds to the noise part.

The experimentally obtained output signal in the k-th order moment is contaminated by the fluctuation at the output channel as $\chi_k s + \eta_k$. Here, the offset m_k is subtracted from the moment M_k. In order to obtain the SNR, it is required to calculate $\{m_k\}$ and $\{\chi_k\}$. These are calculated from the probability density of the noise z_n with the form of the function F. Furthermore, the calculation of the SNR needs the quantity $\langle \eta_k^2\rangle$, since the SNR for the k-th order moment is given as $R_k = \chi_k^2 s^2/\langle\eta_k^2\rangle$. Using the covariance of the fluctuations in the k-th and l-th order moments $\langle\eta_k\eta_l\rangle = V_{kl}/N$, where $V_{kl} \equiv \langle\left[F^k(s+z) - m_k\right]\left[F^l(s+z) - m_l\right]\rangle$, the SNR is expressed as $R_k = s^2 N\hat{R}_k$ with the definition $\hat{R}_k = \chi_k^2/V_{kk}$. In this paper, $\hat{R}_k$ is referred as the *scaled SNR* for the k-th order moment.

Above, a constant input signal has been assumed. However, for the sake of information transmission, temporally varying input should be considered. If the input signal varies slowly, we can indeed calculate the SNR under such a condition. Consider a slowly varying signal s_n: $s_n = \tilde{s}_i$ with $n = Ni + j$ ($j = 0, \cdots, N-1$ and $i = 0, \cdots, \hat{N}/N - 1$). Then the signal s_n is constant in a time scale of N but varies in the larger time scale $\hat{N} \gg N$. In the large time scale, Eq. (2) is replaced by $M_{k,i} = m_k + \chi_k\tilde{s}_i + \eta_{k,i}$, where $M_{k,i}$ and $\eta_{k,i}$ denote the empirical moment and its fluctuation at the coarse-grained time i, respectively. Then, the SNR is given by $R_k = \hat{N}P_s\hat{R}_k$, where P_s is the power of the input signal $P_s = \sum_{n=0}^{\hat{N}-1} s_n^2/\hat{N}$. Note that the conventional SNR R_k^{conv} for the k-th order moment defined by the power spectrum density $\tilde{S}_{s,k}(\omega)/\tilde{S}_{\eta,k}(\omega)$, where $\tilde{S}_{s,k}(\omega)$ denotes the signal power density of the output at the input signal frequency ω and $\tilde{S}_{\eta,k}(\omega)$ is the noise power density, is connected to the above defined SNR R_k via the relation $R_k^{\mathrm{conv}}\tilde{S}_{\eta,k}(\omega) = R_k\int_{-\infty}^{\infty}\tilde{S}_{\eta,k}(\omega')d\omega'$. This relation follows from the result of Parseval's theorem $\left\langle\eta_k^2\right\rangle = \int_{-\infty}^{\infty}\tilde{S}_{\eta,k}(\omega')d\omega'$. Here we have used the assumption that the noise z_n is white.

In experiment, the SNR is attained from the correlation coefficient C_k between the input and the output. The correlation coefficient C_k is defined as

$$C_k = \frac{\sum_{i=0}^{\hat{N}/N-1}(\tilde{s}_i - \bar{s})(M_{k,i} - m_k)}{\sqrt{\sum_{i=0}^{\hat{N}/N-1}(\tilde{s}_i - \bar{s})^2}\sqrt{\sum_{i=0}^{\hat{N}/N-1}(M_{k,i} - m_k)^2}}, \tag{3}$$

where $\bar{s}$ is the time average of the input signal $\bar{s} = \sum_{i=0}^{\hat{N}/N-1}\tilde{s}_i/(\hat{N}/N) = \sum_{n=0}^{\hat{N}-1}s_n/\hat{N}$. If the input signal s_i is sufficiently small ($|s| \ll 1$) and slow, the approximated expression of the correlation coefficient C_k is given as

$$C_k \approx \left(1 + 1/P_s\hat{R}_k\right)^{-1/2}. \tag{4}$$

This expression bridges a gap between the experimentally obtained correlation coefficient and the theoretically predicted SNR. If the input signal, the probability density of the noise, and the function F are all known, the experimental results can be compared with the theory via this relation.

The scaled SNR $\hat{R}_k$ is shown as a function of the standard deviation σ of the noise z_n in the inset of Fig. 2. We have here used a white Gaussian noise with zero mean, i.e., $\langle z_n\rangle = 0$ and $\langle z_n z_m\rangle = \sigma^2\delta_{nm}$. The input signal is assumed to be sinusoidal: $s_n = \epsilon\cos n\omega$ with $\epsilon = 0.1$ and $\omega = 0.001$. We have used a 3-valued function as F: $F(y) = a_0$ for $y < \theta_1$, $F(y) = a_1$ for $\theta_1 \le y < \theta_2$, and $F(y) = a_2$ for $y \ge \theta_2$ with the parameter sets $(a_0, a_1, a_2) = (3.0, -10.0, 1.0)$ and $(\theta_1, \theta_2) = (-1.0, 1.0)$, i.e., $I_0 = (-\infty, -1.0)$, $I_1 = [-1, 1.0)$, $I_2 = [1.0, \infty)$. The theoretical curves are directly given by the relation $\hat{R}_k = \chi_k^2/V_{kk}$, and $\hat{R}_k$ for the simulation results are calculated from the correlation coefficient (4).

Fig. 2 clearly shows that the first and the second moments both exhibit SR, which is characterized by the resonance-like peaks. Note that, in general, the locations and the heights of the peaks for $\hat{R}_k$ both depend on the system settings $\{a_k\}$ and $\{I_k\}$. Furthermore, the occurrence of the SR, i.e., the non-monotonic change of the SNR in each moment also depends on $\{a_k\}$ and $\{I_k\}$.

3. Information Aggregation via PCA

Aggregation of the information in the statistically independent moments can yield the signal detection performance higher than the conventional SR using only the first

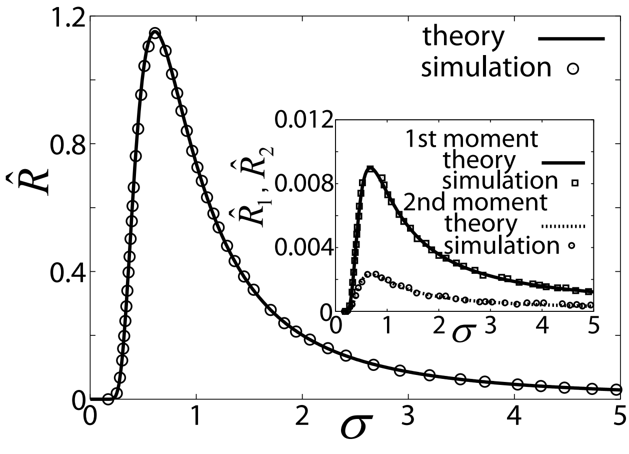

Figure 2: Scaled SNR $\hat{R}$ obtained by our information aggregation method using PCA, and those for the first moment $\hat{R}_1$ and the second moment $\hat{R}_2$. The theoretical results (solid and dashed lines) are directly obtained from Eq. (8) and $\hat{R}_k = \chi_k^2/V_{kk}$. The simulation results (circles and squares) are evaluated from the relation between the correlation coefficients and the SNR Eqs. (4) and (10).

moment. The appropriate aggregation is realized by the statistical tool known as PCA [12, 13]. PCA provides the appropriate linear projection of the vector $(M_1, \cdots, M_{K-1})$ modulated by the input signal in $(K-1)$-dimensional space to elicit the signal part. Such a projection gives the maximum SNR.

Consider a linear projection of $(M_1, \cdots, M_{K-1})$ onto $\mathbb{R}$. This linear projection is expressed as

$$y = \sum_{k=1}^{K-1} w_k M_k, \tag{5}$$

where the weight w_k is determined so as to maximize the SNR as follows. If the input is a constant small signal, the formal expression of the SNR for y is given as

$$R = s^2 N \left(\sum_{k,l=1}^{K-1} w_k w_l \chi_k \chi_l \right) \left(\sum_{k,l=1}^{K-1} w_k w_l V_{kl} \right)^{-1}, \tag{6}$$

because of the asymptotic expression of the projected output y: $y = \sum_{k=1}^{K-1} w_k(m_k + \chi_k s + \eta_k) + O(s^2)$. Then, the maximization condition on the SNR $\partial R/\partial w_k = 0$ for all $k(= 1, \cdots, K-1)$ determines the optimal weight as

$$w_k = \alpha \sum_{l=1}^{K-1} V_{kl}^{-1} \chi_l, \tag{7}$$

where V_{kl}^{-1} is the inverse matrix of the covariance matrix V, and α is an arbitrary (non-zero) real constant. Substituting the optimal weight (7) into Eq. (6), we find the SNR as

$$R = s^2 N \hat{R}, \qquad \hat{R} \equiv \sum_{k,l=1}^{K-1} V_{kl}^{-1} \chi_k \chi_l. \tag{8}$$

For a slowly varying signal, similarly to the replacement in the previous section, the SNR (8) is replaced by $R = \hat{N} P_s \hat{R}$.

The important feature of the SNR obtained by PCA is that it does not depend on the choice of $\{a_k\}$. R is invariant under a non-singular linear transformation of the empirical moments $M_k \to \sum_{l=1}^{K-1} A_{kl} M_l$, where A is an arbitrary regular matrix. This is straightforwardly checked from the transformation of the weight w_k and the covariance matrix V: $w_k \to \sum_{l=1}^{K-1} A_{kl}^{-1} w_l$ and $V_{kl} \to \sum_{k',l'=1}^{K-1} A_{kk'} V_{k'l'} A_{l'l}$. Note also that the empirical probability $P_k = \sum_{n=1}^{N} \delta(x_n, a_k)/N$ is given as a linear combination of the empirical moments. Against this background, the values of $\{a_k\}$ play the role of a label on P_k only. Therefore, the SNR acquired by PCA is concluded to be determined only by the empirical probabilities $\{P_k\}$. In the above expression, the notation of the Kronecker delta $\delta(u, v) = \delta_{uv}$ has been used. In addition, the SNR (8) does not depend on the arbitrary constant α.

In contrast to the invariance on the choice of $\{a_k\}$, the domains $\{I_k\}$ controls the height and the location of the peak of the SNR R. This is because the empirical probability $\{P_k\}$, which is the source of the information on the input signal, depends on $\{I_k\}$. Furthermore, the domains $\{I_k\}$ affect the non-monotonicity of R, i.e., the occurrence of SR.

Because of the optimization, the SNR obtained via PCA gives the highest SNR over those for all possible linear combinations of the empirical moments. Therefore, the SNR acquired in this framework is confirmed to be higher than the SNR for any single moment, i.e., $R \geq R_k$. As a consequence, the information aggregation method using PCA yields the signal detection performance higher than the conventional SR.

Note that, only in a 2-valued system, the conventional SR, which uses only the first moment, is equivalent to our information aggregation method. This is because the first moment is the unique statistically independent moment in a 2-valued system. Note that the conventional SR in a 2-valued system reproduces the same signal detection performance as that is attained by the maximum likelihood method, which is based on the empirical probability P_k [6, 8, 9].

Similarly to the case in the previous section, the input–output correlation coefficient connects experimentally obtained data with the SNR theoretically calculated above. The probability density of the input noise z_n with the form of the function F gives the concrete values of $\{\chi_k\}$ and $\{V_{kl}\}$. Then, the scaled SNR $\hat{R}$ is calculated from Eq. (8). In contrast, in experiments, the decomposition of the output data into the signal and noise parts in order to acquire $\{\chi_k\}$ and $\{V_{kl}\}$ to evaluate the SNR. However, for the purpose of checking the coincidence between the theory and the experiment, such a decomposition of the outputs is not required. The correlation coefficient provides the SNR without the decomposition of the output signal. Under the condition where the probability density of the input noise z_n, the function F, and the slowly varying input signal $\tilde{s}_i$ are all given, the correlation coefficient is calculated from the experimentally obtained data $\{M_{k,i}\}$. The input–output correlation coefficient is defined as

$$C = \frac{\sum_{i=0}^{\hat{N}/N-1} S_i \sum_{k,l=1}^{K-1} \mu_{k,i} V_{kl}^{-1} \chi_l}{\sqrt{\sum_{i=0}^{\hat{N}/N-1} S_i^2} \sqrt{\sum_{i=0}^{\hat{N}/N-1} \left(\sum_{k,l=1}^{K-1} \mu_{k,i} V_{kl}^{-1} \chi_l \right)^2}}, \tag{9}$$

where $S_i \equiv \tilde{s}_i - \bar{s}$ and $\mu_{k,i} \equiv M_{k,i} - m_k$. The same notations as in the case for C_k have been used for the input and output signals. Note that the correlation coefficient C can be expressed asymptotically in terms of the scaled SNR as

$$C \approx \left(1 + 1/P_s\hat{R}\right)^{-1/2}. \tag{10}$$

Since $\{S_i\}$, $\{\mu_{k,j}\}$ and V_{kl} all can be calculated from the given function F and the probability density of the input noise z_n with observed data $\{M_{k,j}\}$, the correlation coefficient C can be obtained experimentally. Then, the scaled SNR $\hat{R}$ is calculated by the above relation.

In this framework, the input signal is estimated as $\hat{s} = \sum_{k=1}^{K-1} w_k(M_k - m_k)/\sum_{k=1}^{K-1} w_k\chi_k$, which is an experimentally obtained random variable. The SNR R gives the accuracy of this estimation.

In Fig. 2, the theoretically obtained scaled SNR is compared with that evaluated from numerical simulations. The function F, the input noise $\{z_n\}$, and the input signal $\{s_n\}$ are the same as those for $\hat{R}_k$ in the inset of Fig. 2. In Fig. 2, we find a resonance-like peak in the SNR attained via PCA that implies the occurrence of SR. Note that the peak height of $\hat{R}$ is significantly larger than that of each moment $\hat{R}_k$. Thus, the signal detection performance is concluded to be improved by the information aggregation via PCA, compared with the conventional SR using only the first moment.

4. Concluding Remarks

In this paper, the novel type of SR phenomena, which occur in moments, is presented. In order to investigate the signal detection performance using such phenomena, we have obtained the formal expression of the SNR for moments. Furthermore, we have proposed the information aggregation method that yields the SNR higher than that for any single moment. The aggregation method is realized by PCA, and collects the information contained in statistically independent moments.

An advantage of our proposed information aggregation method is its high signal detection performance. Our approach is confirmed to yield the signal detection performance higher than that achieved by the conventional SR, which uses only the first moment of the outputs.

Another advantage of our method is its robustness against the change of system parameters. The SNR obtained via our approach is independent of the details of the system parameters $\{a_k\}$. In contrast, the conventional SR is sensitive to the change of such system parameters.

References

[1] R. Benzi, A. Sutera, and A. Vulpiani, "The mechanism of stochastic resonance," *J. Phys. A: Math. Gen.*, vol.14, pp.L453–L457, 1981.

[2] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, "A Theory of Stochastic Resonance in Climatic Change," *Tellus*, vol.34, pp.10–16, 1982.

[3] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, "Stochastic Resonance," *Rev. Mod. Phys.*, vol.70, pp.223–287, 1998.

[4] P. Jung and P. Hänggi, "Amplification of small signals via stochastic resonance," *Phys. Rev. A*, vol.44, 8032–8042, 1991.

[5] S. Fauve and F. Heslot, "Stochastic resonance in a bistable system," *Phys. Lett. A*, vol.97, pp.5–7, 1983.

[6] B. Kosko and S. Mitaim, "Stochastic Resonance in Noisy Threshold Neurons," *Neural Networks*, vol.16, pp.755–761, 2003.

[7] A. D. Hibbs, E. W. Jacobs, A. R. Bulsara, J. J. Bekkedahl, and F. Moss, "Signal enhancement in a r.f. SQUID using stochastic resonance," *Nuovo Cimento D*, vol.17, pp.811–817, 1995.

[8] B. Kosko and S. Mitaim, "Robust stochastic resonance for simple threshold neurons," '*Phys. Rev. E*, vol.70, no.031911, pp.1–10, 2004.

[9] M. D. McDonnell, D. Abbott, C. E. M. Pearce, "An analysis of noise enhanced information transmission in an array of comparators," *Microelectronics Journal*, vol.33, pp.1079–1089, 2002.

[10] F. Chapeau-Blondeau, "Periodic and aperiodic stochastic resonance with output signal-to-noise ratio exceeding that at the input," *Int. J. Bifur. Chaos* vol.9, pp.267–272, 1999.

[11] L. Gammaitoni, "Stochastic resonance and the dithering effect in threshold physical systems," *Phys. Rev. E*, vol.52, pp.4691–4698, 1995.

[12] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol.6, pp.559–571, 1901.

[13] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol.25, pp.417–441, 1933.