A STOCHASTIC EVALUATION OF ACTUAL SOUND ENVIRONMENT BASED ON TWO TYPE INFORMATION PROCESSING METHODS—THE USE OF EXPANSION SERIES TYPE REGRESSION AND FUZZY PROBABILITY

Akira IKUTA and Mitsuo OHTA

Faculty of Engineering, Kinki University,
Takaya, Higashihiroshima city, 729-17 Japan.

ABSTRACT: In the actual sound environment, the random signal often shows a complex fluctuation pattern apart from a standard Gaussian distribution. In this study, an evaluation method for the sound environmnetal system is proposed in the generalized form applicable to the actual stochastic phenomena, by introducing two type information processing methods based on the regression model of expansion series type and the Fuzzy probability. The effectiveness of the proposed method are confirmed experimentally too by applying it to the observed data in the actual noise environment.

1. INTRODUCTION

The random signal in the actual sound environment usually exhibits multifarious and complex characteristics such as non-Gaussian distribution, non-linearity and non-stationary properties, owing to natural, social and/or human factors. Furthermore, the observation data are often contaminated by the background noise of arbitrary distribution type. In this study, an evaluation method for the sound environmental system is proposed in the generalized form applicable to the actual stochastic phenomena, by introducing two types of information processing.

More specifically, first, the well-known standard type ideal cases based on Gaussian, linear and stationary properties are adopted as the basis of the present theoretical expression. Next, by utilizing positively the general extension capability of two type[1] information processing methods based on regression models of the expansion series type and using the Fuzzy probability[2] the above ideal cases can be extended to the form applicable to more complicated cases with non-Gaussian, non-linear and non-stationary properties. In particular, by employing the above expansion series type regression models, one can successively investigate the diversity and complexity of stochastic properties step by step, on the basis of employing the lower order information. Here, the higher order concepts of statistical properties are reflected in each expansion coefficients of the expansion series expression. In the first part of this study, a digital filter for estimating recursively the fluctuation wave form of only the latent specific noise based on the observed data contaminated by a background noise of non-Gaussian type is considered. Based on the observed noisy data with the background noise, in order to estimate several evaluation quantities for the above specific noise (e.g., Lx, Leq and peak value, etc.), it is fundamental to estimate the fluctuation wave form of only the specific noise at every instantaneous time. In this study, a state estimation method for the specific noise is derived by adopting the whole of conditional probability density function (abbr. pdf.) of a time series regression type in the expansion series form, as the system transition law. On the other hand, in the case of applying the Fuzzy probability to an idealized standard model, the general extension capability to the actual complicated situation can be achieved by estimating the parameters of membership function based on the actually observed data with non-Gaussian type fluctuation. Concretely, in the second part of this study, for the purpose of finding out a practical method, the simplified expression with only a few parameter is adopted as the basis of theoretical expression. Next, by utilizing positively the generalized capability existing in the information processing of Fuzzy probability, this ideal model can be extended to the form applicable to more complicated general cases. Finally, the effectiveness of the proposed method are confirmed through only the principle experiments by applying it to the observed data in the real noise environment.

2. A GENERALIZED REGRESSION ANALYSIS OF EXPANSION SERIES TYPE[1]

In the case of paying our attention to a prediction variable x and a criterion variable y, every information on linear and/or nonlinear correlations between x and y is originally included in the conditional pdf. $P(y|x)$. Especially, as a typical regression relationship between the above two variables, the following regression function (i.e., a conditional expectation of y) can be adopted as follows:

$$y(x)=\int y\, P(y|x)\, dy. \tag{1}$$

In order to explicitly find various correlation properties between x and y, let us expand the joint pdf. $P(x, y)$ into an orthogonal polynomial series as follows:

$$P(x, y)=P(x)P(y)\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} A_{mn}\, \psi_m^{(1)}(x)\, \psi_n^{(2)}(y), \quad A_{mn}=\langle\psi_m^{(1)}(x)\, \psi_n^{(2)}(y)\rangle, \tag{2}$$

where $P(x)$ and $P(y)$ are pdfs. of x and y. Two functions $\psi_m^{(1)}(x)$ and $\psi_n^{(2)}(y)$ are orthonormal polynomials with the weighting functions $P(x)$ and $P(y)$. The information on various types of linear and/or nonlinear correlations between x and y is reflected hierarchically in each expansion coefficient A_{mn}. By using Eq.(2) and the well-known Bayes' theorem, the conditional pdf. having the whole information on the regression relationship can be derived as

$$P(y|x)=P(x, y)/P(x)=P(y)\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} A_{mn}\, \psi_m^{(1)}(x)\, \psi_n^{(2)}(y). \tag{3}$$

Therefore, by considering the orthonormal condition of $\psi_n^{(2)}(y)$, the regression relationship defined by Eq.(1) can be concretely given in the expansion series form, as follows:

$$y(x)=\sum_{m=0}^{\infty}\sum_{n=0}^{1} A_{mn}\, \psi_m^{(1)}(x)\, C_n, \quad (y=\sum_{j=0}^{1} C_j\, \psi_j^{(2)}(y),\ C_j\text{: constants determined in advance}). \tag{4}$$

In the case when the expansion coefficients A_{mn} can be considered as an invariant characteristic reflecting the correlation relationship between two variables x and y, the output response of the system for the arbitrary other kinds of input can be predicted by replacing x in Eq.(4) with the arbitrary input. Furthermore, based on the conditional pdf. of Eq.(3), the output response probability distribution for the arbitrary random excitation can be estimated as $P_s(y)=\langle P(y|x)\rangle_x$.

3. DIGITAL FILTER FOR ESTIMATING SPECIFIC NOISE BASED ON EXPANSION SERIES TYPE REGRESSION

3.1 Stochastic Model for Specific Noise under Existence of Background Noise

Let the specific noise power at a discrete time k be x_k. In this section, a digital filter for estimating the latent specific noise under the existence of a background noise is proposed. For the derivation of the digital filter, a dynamical model on the specific noise must be first established. Here, a system model in the form of transition probability instead of the usual simplified system models is adopted as one of stochastic dynamical models. Concretely, by replacing x and y in Eq.(3) with x_k and x_{k+1}, the conditional probability density function $P(x_{k+1}|x_k)$ reflecting all linear and nonlinear correlation information on a time transition from x_k to x_{k+1} of the state is expressed as

$$P(x_{k+1}|x_k)=P(x_{k+1})\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}A_{mn}\,\psi_m(x_{k+1})\,\psi_n(x_k),\quad A_{mn}=<\psi_m(x_{k+1})\,\psi_n(x_k)>. \tag{5}$$

On the other hand, by using the additive property of acoustic power, the observation power y_k under the existence of background noise can be expressed as follows:

$$y_k=x_k+v_k. \tag{6}$$

We assume that the statistics of background noise power v_k are known in advance. Each expansion coefficient A_{mn} in Eq.(5) must be estimated based on the noisy observation y_k because the instantaneous values of x_k are unknown. After paying our attention to the correlation information between y_{k+1} and y_k, applying the binomial theorem and using the statistical independency between x_k and v_k, the following relationship can be derived.

$$<y_{k+1}^m\,y_k^n>=<(x_{k+1}+v_{k+1})^m\,(x_k+v_k)^n>=\sum_{i=0}^{m}\sum_{j=0}^{n}a_{mi}\,b_{nj}\,A_{ij}<v_{k+1}^{m-i}\,v_k^{n-j}>, \tag{7}$$

where a_{mi} and b_{nj} are appropriate constants determined in advance. By solving Eq.(7) inversely in the successive relationship starting from the lower order statistics (i.e., m, n=0, 1, 2, ...), the expansion coefficients A_{mn} can be evaluated from the correlation quantities $<y_{k+1}^m\,y_k^n>$ and the statistics of the background noise.

3.2 Derivation of Estimation Algorithm

Based on the Bayes' theorem, the estimate of an arbitrary polynomial function $f_L(x_k)$ of x_k with L-th order can be derived in an infinite series expression[3] as follows:

$$\widehat{f_L}(x_k)=<f_L(x_k)|Y_k>=\sum_{r=0}^{L}\sum_{s=0}^{\infty}d_{Lr}\,B_{rs}\,\theta_s^{(2)}(y_k)\Big/\sum_{s=0}^{\infty}B_{0s}\,\theta_s^{(2)}(y_k),\quad \Big(f_L(x_k)=\sum_{r=0}^{L}d_{Lr}\,\theta_r^{(1)}(x_k)\Big) \tag{8}$$

with

$$B_{rs}=<\theta_r^{(1)}(x_k)\,\theta_s^{(2)}(y_k)|Y_{k-1}>, \tag{9}$$

where $Y_k(=\{y_1, y_2, ..., y_k\})$ is a set of observation data until a time k and the coefficients d_{Lr} are appropriate constants. Two functions $\theta_r^{(1)}(x_k)$ and $\theta_s^{(2)}(y_k)$ are the orthonormal polynomials of degrees r and s, with the weighting functions $P_0(x_k|Y_{k-1})$ and $P_0(y_k|Y_{k-1})$ which can be artificially chosen as the pdfs. describing the dominant parts of the actual fluctuation, or as the well-known standard pdfs. like the Gaussian or Gamma distribution functions. The expansion coefficient B_{rs} in Eq.(9) can be given by the statistics of the background noise v_k and the prediction of x_k at a discrete time k-1 (i.e., the expectation for arbitrary functions of x_k conditioned by Y_{k-1}). Next, let the arbitrary polynomial function with the L-th order of x_{k+1} be $g_L(x_{k+1})$. Using the property of conditional expectation, the prediction algorithm for the function $g_L(x_{k+1})$ at a discrete time k can be given by

$$g_L^*(x_{k+1})=<g_L(x_{k+1})|Y_k>=<<g_L(x_{k+1})|x_k, Y_k>|Y_k>=<\int g_L(x_{k+1})P(x_{k+1}|x_k)dx_{k+1}|Y_k>. \quad (10)$$

The last equality of Eq.(10) is approved by the natural inference that if the statistical relation between x_k and x_{k+1} can be sufficiently represented by Eq.(5), the remaining fluctuation factor is an accidental error and is independent of Y_k. After expanding the function $g_L(x_{k+1})$ into an orthogonal series with the expansion coefficients e_{Lm}, considering the orthonormal condition of the polynomial $\psi_m(x_{k+1})$, the expression:

$$g_L^*(x_{k+1})=\sum_{m=0}^{L}\sum_{n=0}^{\infty}e_{Lm}\, A_{mn} <\psi_n(x_k)|y_k>, \quad (11)$$

can be derived. Equation (11) means that the prediction at a discrete time k is expressed in the combination form of estimates for the polynomial function of x_k. Therefore, by combining the estimation algorithm with the prediction algorithm, the recurrence estimation of the latent specific noise can be achieved.

4. EVALUATION METHOD FOR SOUND ENVIRONMENT BASED ON FUZZY PROBABILITY

4.1 Utilization of Fuzzy Moment

The regression parameter (i.e., expansion coefficients A_{mn}) in the regression function of Eq.(4) can be estimated by use of the moment method based on the simultaneous observation data of x and y according to the definition of A_{mn}. However, since there exist partly the observed data with the lower reliability, it is often effective to utilize the observation data in a form that the higher weight is imposed on the data with higher reliability. For the estimation error: $\varepsilon=y-\hat{y}(x)$, applying the well-known least squares method, only a few regression parameter A_{mn} ($m\le M$, $n\le N$) with the lower order is estimated. Hereupon, the larger the value of ε becomes, the more unreliable the data become, because of the increase of the fuzziness on information. Therefore, by utilizing the generalized capability existing in Fuzzy probability[2),]in order to compensate the truncation of the higher expansion terms and the observation data with the lower degrees of reliability, a new criterion considering a weight for the mean squares error is introduced as follows:

$$J_1=\int\varepsilon^2\,\mu_\varepsilon(\varepsilon)\,P(\varepsilon)\,d\varepsilon=<\varepsilon^2\,\mu_\varepsilon(\varepsilon)>, \quad (12)$$

where $\mu_\varepsilon(\varepsilon)$ is a membership function describing the degree of reliability for ε. For example, the following function can be adopted:

$$\mu_\varepsilon(\varepsilon)=1/\{1+(\varepsilon/b)^2\}. \quad (13)$$

Though the parameter b in Eq.(13) can be generally given based on the prior information (or, through trial and error), it can be regarded as an unknown parameter and is estimated simultaneously with the regression parameters A_{mn}. Accordingly, the parameters A_{mn} and b are determined so as to satisfy the relationships: $\partial J_1/\partial A_{mn}=0$, $\partial J_1/\partial b=0$.

4.2 Utilization of Fuzzy Entropy

In the case of predicting the output y for an arbitrary input x by use of Eq.(4), it is assumed that the linear and/or nonlinear correlation informations between x and y give an invariant characteristic expressing the statistical relationship between the input and output. However, strictly speaking, especially in a nonlinear system, the regression parameters depend on the fluctuation range of the input. In this section, an estimation method for the regression parameters reflecting the change of the statistical properties of input is proposed by introducing the Fuzzy entropy[2]. Concretely, under the assumption that the mean x^* and the standard deviation $\sigma_x{}^*$ of the arbitrary input in the case of prediction are known in advance, the regression parameters are estimated by use of a criterion reflecting these two statistics on the input. First, let us pay our attention to the conditional entropy:

$$I(y|x)=-\iint P(x, y) \log P(y|x)\, dxdy. \tag{14}$$

Next, by introducing a membership function:

$$\mu_x(x;\, x^*, \sigma_x{}^*, \mathbf{a})=1\ (0<|X|<X_1),\ a_i\ (X_i<|X|<X_{i+1};\ i=1,2,\ldots,I),\ 0\ (X_{I+1}<|X|),$$
$$X=(x-x^*)/\sigma_x{}^*,\ x^*=\langle x\rangle,\ \sigma_x{}^*=\langle(x-x^*)^2\rangle,\ \mathbf{a}=(a_1, a_2, \ldots, a_I)\ (0<a_I<\cdots<a_2<a_1), \tag{15}$$

reflecting two statistics x^* and $\sigma_x{}^*$ of the input in the prediction step, the regression parameters can be estimated so as to minimize the Fuzzy entropy:

$$H(y|x)=-\iint \mu_x(x;\, x^*, \sigma_x{}^*, \mathbf{a})\, \mu_y(y)\, P(x, y) \log P(y|x)\, dxdy. \tag{16}$$

The membership function in Eq.(15) expresses the degree of confidence of observation data. Since the statistics of the output are unknown, we adopt the relation: $\mu_y(y)=1$ $(-\infty<y<\infty)$ in Eq.(16). Therefore, the parameters A_{mn} and $\mathbf{a}$ can be estimated so as to maximize the function: $J_2=\langle\mu_x(x;\, x^*, \sigma_x{}^*, \mathbf{a}) \log P(y|x)\rangle$.

5. APPLICATION TO OBSERVED DATA IN NOISE ENVIRONMENT

5.1 Experiment for Digital Filter Based on Expansion Series Type Regression

In order to examine the practical usefulness of the digital filter for state estimation in Sect. 3, the method is applied to the road traffic noise data. Applying the proposed estimation method to the actually observed data contaminated by the background noise, the fluctuation wave form of the latent specific noise is estimated. Figure 1 shows one of the estimation results. For comparison, the estimation result calculated based on the usual linear system model is also shown in this figure. There are great discrepancies between the estimates based on the

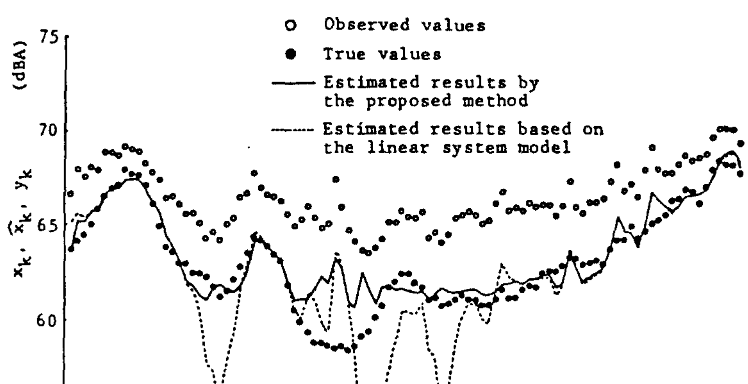

Fig.1 State estimation results for the road traffic noise contaminated by a background noise.

linear system model and the true values, while the proposed method is estimated precisely the wave form of the road traffic noise with rapidly changing fluctuation.

5.2 Experiment for Noise Evaluation Method Based on Fuzzy Probability

The proposed theory in Sect. 4 has been applied to the actual data observed in a complicated sound environmental system which can not be identified from the standard structural approach. After regarding the sound insulation system as a single-input – single-output system for the simplification of procedure, the regression relationship between the input and the output variables have been evaluated. Next, by use of the obtained regression relationships, the output response probability distribution excited by an arbitrary another input signal is predicted. Figure 2 shows the predicted result of the output probability distribution by applying the method based on Fuzzy entropy. The theoretically predicted curve based on the Fuzzy entropy shows better agreement with the experimentally sampled values than the result without considering Fuzzy probability.

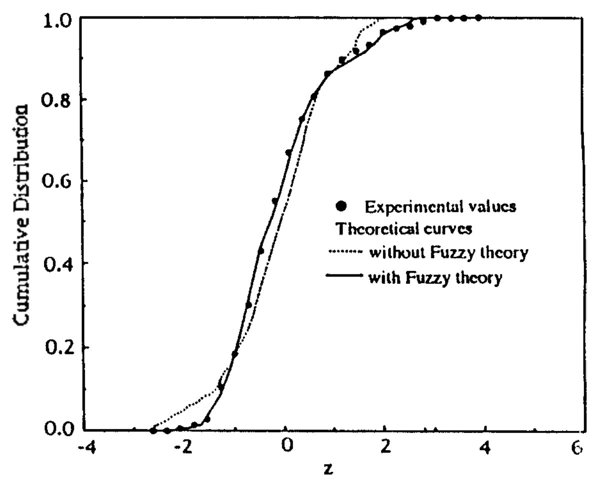

Fig.2 Comparison between theoretically predicted curves by use of Fuzzy entropy and experimentally sampled points on the output probability distribution.

6. CONCLUSION

In this study, a stochastic evaluation method for a complicated sound environment which usually shows non-linear, non-stationary and non-Gaussian properties has been proposed. The regression models using an expansion series type and the well-known Fuzzy probability have enabled us to treat the above complicated phenomana by adopting some simplified standard expressions of pdf. with only a few parameters. The principle validity of these two proposed methods has been confirmed by employing the actually observed data in noise environment.

ACKNOWLEDGMENTS
We would like to express our cordial thanks to Prof. K. Hatakeyama, Dr. N. Nakasako, Messrs. M. Ishimoto and K. Yamamoto for their helpful assistance.

REFERENCES

1. M. Ohta, A. Ikuta and N. Takaki, "A new trial of regression analysis with less information loss and its application to sound insulation walls—Matching to the prediction of response probability distribution," J. Acousti. Soc. Jpn. 44, 848-856 (1988).
2. L. A. Zadeh, "Probability measures of fuzzy events," J. Math. Anal. Appl. 28, 421-427 (1968).
3. M. Ohta, K. Hatakeyama and M. Nishimura, "Some unified methods in multivariable linear filtering and prediction problems by use of the expansion form of Bayes' theorem," Proc. of the 26th Japan National Congress for App. Mech. 26, 383-398 (1976).