LA-UR-02-1240: SPIE's 9th Annual International Symposium on Smart Structures and Materials, San Diego, CA, March 17-21, 2002.

Consideration of Environmental and Operational Variability for Damage Diagnosis

Hoon Sohn[a], Keith Worden[b], and Charles R. Farrar[a]

[a] Los Alamos National Laboratory, MS P946, Los Alamos, NM 87545, USA.
[b] Department of Mechanical Engineering, University of Sheffield, Sheffield, UK.

ABSTRACT

Damage diagnosis is a problem that can be addressed at many levels. Stated in its most basic form, the objective is to ascertain simply if damage is present or not. In a statistical pattern recognition paradigm of this problem, the philosophy is to collect baseline signatures from a system to be monitored and to compare subsequent data to see if the new "pattern" deviates significantly from the baseline data. Unfortunately, matters are seldom as simple as this. In reality, structures will be subjected to changing environmental and operational conditions that will affect measured signals. In this case, there may be a wide range of normal conditions, and it is clearly undesirable to signal damage simply because of a change in the environment. In this paper, a unique combination of time series analysis, neural networks, and statistical inference techniques is developed for damage classification explicitly taking into account these natural variations of the system in order to minimize false positive indication of true system changes.

Keywords: Damage classification, time series analysis, auto-associative neural networks, statistical inference, operational and environmental variations.

1. INTRODUCTION

Structural health monitoring is a problem, which can be addressed at many levels. Stated in its most basic form, the objective is to ascertain simply if damage is present or not. The philosophy is simple: During the normal operation of a system or structure, measurements are recorded and features are extracted from data, which characterize the normal conditions. After training the diagnostic procedure in question, subsequent data can be examined to see if the features deviate significantly from the norm. That is, a simple damage classifier such as outlier analysis (Worden, 1997; Worden et al., 2000) can be employed for deciding if measurements from a system or structure indicate significant departure form previously established normal conditions. An alarm is signaled if observations increase above a pre-determined threshold.

Unfortunately, matters are seldom as simple as this. In reality, structures will be subjected to changing environmental and operational states such as varying temperature, moisture, and loading conditions affecting the measured features and the normal condition. In this case, there may be a continuous range of normal conditions, and it is clearly undesirable for the damage classifier to signal damage simply because of changes in the environment or operation. In fact, these changes can often mask subtler structural changes caused by damage (Sohn et al., 2001a).

Data normalization is a procedure to "normalize" data sets such that signal changes caused by operational and environmental variations of the system can be separated from structural changes of interests, such as structural deterioration or degradation. One approach to solving this problem is to measure parameters related to these environmental and operational conditions as well as the vibration features over a wide range of these varying conditions to characterize the normal conditions. The normal conditions can be then parameterized to reflect the different environmental and operational states. Such a parameterization study is applied to vibration signals obtained from the Alamosa Canyon Bridge in New Mexico, USA to relate the change of the bridge's fundamental frequency to the temperature gradient of the bridge (Sohn et al, 1999). The measured fundamental frequency of the Alamosa Canyon Bridge in New Mexico varied approximately 5% during a 24-hour test period, and the change of the fundamental frequency was well correlated to the temperature difference across the bridge deck. Because the bridge is approximately aligned in the north and south direction, there is a large temperature gradient between the west and east sides of the bridge deck throughout the day. A damage classifier, which does not provide false indication of damage under changing environmental and operational conditions, can be then built. On the other hand,

[a] Correspondence to Hoon Sohn, Email: sohn@lanl.gov, Tel: 505-663-5205, Fax: 505-663-5225.

there are cases where it is difficult to measure parameters related to the environmental and operational conditions. Furthermore, if damage produces a change in the measured signals that is in someway orthogonal or uncorrelated to the change caused by the environmental or operational variability, it may be possible to distinguish the change in the measured signals caused by damage from that caused by the sources of variability without a measure of the operational or environmental variability. This paper addresses the later cases where no measurements are available for these natural variations. Other applications of this data normalization are presented in Sohn and Farrar (2001) and Sohn et al. (2001b).

In this paper, a unique combination of time series analysis, auto-associative neural networks, and statistical pattern recognition techniques is developed to automate a damage identification problem with a special attention to data normalization. First, a time prediction model, called an Auto Regressive-Auto Regressive model with Exogenous inputs (AR-ARX) model is fit to vibration signals measured during normal operating conditions of the structure. Next, data normalization is performed based on the auto-associative neural network where target outputs are simply inputs to the network. Using the extracted features, which are the parameters of the AR-ARX model, corresponding to the normal conditions as inputs, the auto-associative neural network is trained to characterize the underlying dependency of the extracted features on the unmeasured environmental and operational variations by treating these environmental and operational conditions as hidden intrinsic variables in the neural network. When a new time signal is recorded from an unknown state of the system, the parameters of the time prediction model are computed for the new data set and are fed to the trained neural network. When the structure undergoes structural degradation, it is expected that the prediction errors of the neural network will increase for the damage case. Based on this premise, a damage classifier is constructed using a hypothesis testing technique called a sequential probability ratio test (SPRT). The SPRT is one form of parametric statistical inference tests and the adoption of the SPRT to damage detection problems can improve the early identification of conditions that could lead to performance degradation and safety concerns.

The layout of this paper is as follows. Section 2 briefly reviews the time series analysis of vibration signals using the AR-ARX model. In Section 3, a description of the auto-associative neural network is given relating this network with Principal Component Analysis (PCA) and Nonlinear Principal Component Analysis (NLPCA). Section 4 outlines the main theory of the sequential probability ratio test (SRPT), and the proposed approach is applied to experimental data obtained from an eight degree-of-freedom (DOF) system in Section 5. Section 6 concludes and summarizes the findings of this study.

2. TIME SERIES ANALYSIS

A linear prediction model combining AR and ARX models is employed to compute input parameters for the subsequent analysis of an auto-associative neural network presented in Section 3. First, all time signals are standardized prior to fitting an AR model such that:

$$\hat{x} = \frac{x - \mu_x}{\sigma_x} \tag{1}$$

where $\hat{x}$ is the standardized signal, μ_x and σ_x are the mean and standard deviation of x, respectively. This standardization procedure is applied to all signals employed in this study. (However, for simplicity, x is used to denote $\hat{x}$ hereafter.)
For a given time signal $x(t)$, an AR model with r auto-regressive terms is constructed. An AR(r) model can be written as (Box et al., 1994):

$$x(t) = \sum_{j=1}^{r} \phi_{xj}\ x(t-j)\ + e_x(t) \tag{2}$$

The AR order is set to be 30 for the experimental study presented in Section 5 based on a partial auto-correlation analysis described in Box et al. (1994). For the construction of a two-stage prediction model proposed in this study, it is assumed that the error between the measurement and the prediction obtained by the AR model [$e_x(t)$ in Equation (2)] is mainly caused by the unknown external input. Based on this assumption, an ARX model is employed to reconstruct the input/output relationship between $e_x(t)$ and $x(t)$:

$$x(t) = \sum_{i=1}^{p} \alpha_i\ x(t-i)\ + \sum_{j=1}^{q} \beta_j\ e_x(t-j) + \varepsilon_x(t) \tag{3}$$

where $\varepsilon_x(t)$ is the residual error after fitting the ARX(p,q) model to $e_x(t)$ and $x(t)$ pair. The feature for damage diagnosis will later be related to this quantity, $\varepsilon_x(t)$. Note that this AR-ARX modeling is similar to a linear approximation method of an Auto-Regressive Moving-Average (ARMA) model presented in Ljung, 1999 and references therein. Ljung (1999) suggests keeping the sum of p and q smaller than r ($p + q \le r$). Although the p and q values of the ARX model are set rather

arbitrarily, similar results are obtained for different combinations of p and q values as long as the sum of p and q is kept smaller than r. The α_i and β_j coefficients of the ARX model are used as input parameters for the following analysis of the auto-associative neural network. ARX(5,5) is used for this specific experimental study.

3. AUTO-ASSOCIATIVE NEURAL NETWORKS

PCA has been proven to facilitate many types of multivariate data analysis including data reduction and visualization, data validation, fault detection, and correlation analysis (Fukunaga and Koontz, 1970). Similar to PCA, NLPCA is used as an aid to multivariate data analysis. While PCA is restricted on mapping only linear correlations among variables, NLPCA can reveal the nonlinear correlations presented in data. If nonlinear correlations exist among variables in the original data, NLPCA can reproduce the original data with greater accuracy and/or with fewer factors than PCA. This NLPCA can be realized by training a feedforward neural network to perform the identity mapping, where the network outputs are simply the reproduction of network inputs. For this reason, this special kind of neural network is named as an *auto-associative neural network* (Figure 1). The network consists of an internal "bottleneck" layer, two additional hidden layers, and one output layer. The bottleneck layer contains fewer nodes than input or output layers forcing the network to develop a compact representation of the input data. The NLPCA presented in this paper is a general purpose feature extraction/data reduction algorithm discovering features that contain the maximum amount of information from the original data set. In this section, PCA and NLPCA are briefly reviewed. More detailed discussions on PCA, NLPCA, and auto-associative networks can be found from Fukunaga (1990), Kramer (1991), Rumelhart and McClelland (1988), respectively.

3.1. Principal Component Analysis (PCA)

PCA is a linear transformation mapping multidimensional data into lower dimensions with minimum loss of information. Let $\mathbf{Y}$ represent the original data with the size of $m \times l$. Here, m is the number of variables and l is the number of data set. PCA can be viewed as a linear mapping of data from the original dimension m to a lower dimension d:

$$\mathbf{X} = \mathbf{T}\mathbf{Y} \tag{4}$$

where $\mathbf{X}$ ($\in \Re^{d \times l}$) is called the *scores* matrix. $\mathbf{T}$ ($\in \Re^{d \times m}$) is called the *loading* matrix and $\mathbf{T}\mathbf{T}^{\mathrm{T}} = \mathbf{I}$. The loss of information in this mapping can be assessed by re-mapping the projected data back to the original space:

$$\hat{\mathbf{Y}} = \mathbf{T}^{\mathrm{T}}\mathbf{X} \tag{5}$$

Then, the reconstruction error (residual error) matrix E is defined as:

$$\mathbf{E} = \mathbf{Y} - \hat{\mathbf{Y}} \tag{6}$$

The smaller the dimension of the projected space, the greater the resulting error. The loading matrix $\mathbf{T}$ can be found such that the Euclidean norm of the residual matrix, $\|\mathbf{E}\|$, is minimized for the given size of d. It can be shown that the columns of $\mathbf{T}$ are the eigenvectors corresponding to the d largest eigenvalues of the covariance matrix of $\mathbf{Y}$ (Fukunaga, 1990).

3.2. Nonlinear Principal Component Analysis (NLPCA)

NLPCA generalizes the linear mapping by allowing arbitrary nonlinear functionalities. Similar to Equation (4), NLPCA seeks a mapping in the following form:

$$\mathbf{X} = \mathbf{G}(\mathbf{Y}) \tag{7}$$

where $\mathbf{G}$ is a nonlinear vector function and consists of d number of individual nonlinear functions: $\mathbf{G} = \{G_1, G_2, ..., G_d\}$. By analogy to Equation (5), the inverse transformation, restoring the original dimensionality of the data, is implemented by a second nonlinear vector function $\mathbf{H}$:

$$\hat{\mathbf{Y}} = \mathbf{H}(\mathbf{X}) \tag{8}$$

The information lost is again measured by $\mathbf{E} = \mathbf{Y} - \hat{\mathbf{Y}}$. Similar to PCA, $\mathbf{G}$ and $\mathbf{H}$ are computed to minimize the Euclidean norm of $\|\mathbf{E}\|$, meaning minimum information loss in the same sense as PCA. NLPCA employs artificial neural networks to generate these arbitrary nonlinear functions. Cybenko (1989) has shown that functions of the following form are capable of fitting any nonlinear function $\mathbf{y} = g(\mathbf{x})$ to an arbitrary degree of precision:

$$y_k = \sum_{j=1}^{N_2} w_{jk}^2 \, h\left(\sum_{i=1}^{N_1} w_{ij}^1 x_i + b_j \right) \tag{9}$$

where y_k and x_i are the *k*th and *i*th components of **y** and **x**, respectively. w_{ij}^k represents the weight connecting the *i*th node in the *k*th layer to the *j*th node in the *(k+1)*th layer, and b_j is a node bias. N_i is the number of nodes in each layer. $h(x)$ is a monotonically increasing continuous function with the output range of 0 to 1 for an arbitrary input *x*. A sigmoid transfer function is often used in neural networks to realize this function. Note that, to fit an arbitrary nonlinear function, at least two layers of weighted connections are required, and the first hidden layer should be composed of nonlinear transfer functions such as the sigmoid function. Therefore, the two nonlinear vector functions in Equations (7) and (8) should have the same architecture: one hidden layer with nonlinear transfer functions and one output layer. The output layer can have either linear or nonlinear transfer functions without affecting the generality of the mapping.

Now, an auto-associative neural network is constructed by combining mapping **G** and de-mapping **H** functions together as shown in Figure 1. The combined network contains three hidden layers; the mapping, the bottleneck, and de-mapping layers. The second hidden layer is referred to as the *bottleneck layer* because it has the smallest dimension among the three layers. =For instance, the first hidden layer of **G**, which consists of M_1 nodes with nonlinear transfer functions, operates on the columns of **Y** mapping *m* inputs to M_1 node outputs. The output of the first hidden layer is projected into the bottleneck layer, which contains *d* nodes. In a similar fashion, the inverse mapping function **H** takes the columns of **X** as inputs relating *d* inputs to M_2 node outputs. The final output layer reconstructs the target output $\hat{\mathbf{Y}}$, and contains *m* nodes. It should be noted that if the neural networks for **G** and **H** are to be trained separately, the target output **X** is unknown for the training of the **G** network. For the same reason, the input for the **H** network is not known. It is observed that **X** is both the output of **G** and the input of **H**. Therefore, combining the two networks in series, where **G** feeds directly into **H**, results in a new network whose inputs and target outputs are not only known but also identical. Now, the supervised training can be applied to the combined network. Note that the nodes in the mapping and de-mapping layers must have nonlinear transfer functions to model arbitrary **G** and **H** functions. However, nonlinear transfer functions are not necessary in the bottleneck layer. If the mapping and de-mapping layers were eliminated and only the linear bottleneck layer were left, this network would reduce to linear PCA as demonstrated by Sanger (1989). Typically M_1 and M_2 are selected to be larger than *m* and they are set to be equal ($M_1 = M_2$).

In this study, the auto-associative network is employed to reveal the latent relationship between the extracted features and the unmeasured intrinsic parameters causing the variations of the features. Particularly, the auto-associative neural network presented here uses the coefficients of the AR-ARX model presented in the previous section as inputs as well as target outputs, and the network is trained to reveal the inherent excitation level driving the changes. If the neural network is trained to capture the embedded relationships, the prediction error of the neural network will grow when an irrelevant data set, such as ones obtained from a damage state of the system, is fed to the network. Based on this assumption, the auto-associative network is incorporated with the SPRT described in the following section to identify damage.

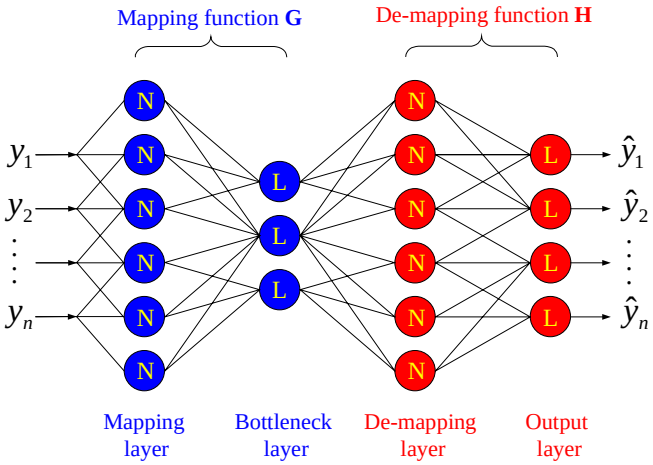

Figure 1: A schematic presentation of an auto-associative neural network

4. SEQUENTIAL PROBABILITY RATIO TEST

In the previous section, the auto-associative neural network is trained using the AR-ARX coefficients as inputs as well as outputs. If $\hat{\alpha}_i$ and $\hat{\beta}_j$ are defined as the outputs estimated from the network, the residual errors using these estimated AR-ARX coefficients can be computed:

$$\varepsilon_y(t) = x(t) - \sum_{i=1}^{p} \hat{\alpha}_i \; x(t-i) - \sum_{j=1}^{q} \hat{\beta}_j \; e_x(t-j) \tag{10}$$

where $\varepsilon_y(t)$ is the residual error obtained using the time series $x(t)$ and the $\hat{\alpha}_i$ and $\hat{\beta}_j$ coefficients estimated from the network. Here, the subscript "*y*" is used to distinguish this residual error from the one shown in Equation (3). When a new set of the AR-ARX coefficients are obtained from a damaged structure and fed to the network, the auto-associative network trained with the undamaged cases will not be able to properly reproduce the new AR-ARX coefficients. Therefore, the standard deviation of the residual error $\varepsilon_y(t)$ associated with the $\hat{\alpha}_i$ and $\hat{\beta}_j$ coefficients is expected to increase compared to that of the baseline residual error $\varepsilon_x(t)$.

Based on this premise, a simple hypothesis test discriminating two hypotheses is constructed using the standard deviation of the residual errors as the parameter in question (Allen et al, 2002):

$$H_o : \sigma(\varepsilon_y) \le \sigma_o, \quad H_1 : \sigma(\varepsilon_y) \ge \sigma_1, \quad 0 < \sigma_o < \sigma_1 < \infty \tag{11}$$

When the standard deviation of the residual error, $\sigma(\varepsilon_y)$, is less than a user specified lower bound σ_o, the system in question is considered undamaged. On the other hand, when $\sigma(\varepsilon_y)$ becomes equal to or larger than the other user specified upper bound σ_1, the system is suspected to be damaged. It should be noted that the selection of σ_o and σ_1 is structure-dependent, and it might be necessary to use signals from a few damage cases as well as from undamaged cases in order to establish these two decision boundaries.

A SPRT starts with observing a sequence of the residual errors, $\{\varepsilon_y(i)\}$ $(i = 1, 2, \ldots)$. This accumulated data set at stage n is denoted as (Ghosh, 1970):

$$E_n = [\varepsilon_y(1), \ldots \varepsilon_y(n)] \tag{12}$$

The goal of a statistical inference is to reveal the probability model of E_n, which is assumed to be at least partially unknown. When the statistical inference is cast as a parametric problem, the functional form of E_n is assumed known and the statistical inference poses some questions regarding the parameters of the probability model. For instance, if $\{\varepsilon_y(i)\}$ are independent and identically distributed (i.i.d.) normal variables, one may pose some statistical test about the mean and/or the variance of this normal distribution.

A sequential test is one of the simplest tests for such a statistical inference where the number of samples required before reaching a decision is not determined in advance. An advantage of the sequential test is that, on average, a smaller number of observations are needed to make a decision compared to the conventional fixed-sample size test. For the well-established fixed-sampling tests, the sample size n is fixed, and an upper bound on the type I error is pre-specified. Then, an optimal fixed-sample test is selected by minimizing the probability of type II error. On the other hand, a sequential test specifies upper bounds on the probabilities of type I and II errors, and minimizes the sample number required to make a decsion. A *type I error* arises if H_o is rejected when in fact it is true. *Type II errors* arise if H_o is accepted when it is false. Among various valid sequential tests, it can be proven that the SPRT minimizes on average the sample size required to make a correction making it an optimal sequential test (Ghosh, 1970). Because of this extreme sensitivity of the SPRT to signal disturbance, the SPRT has been applied for the surveillance of nuclear power plant components (Humenik and Gross, 1991; Cross and Humenik, 1990).

For the hypothesis test in Equation (11), a SPRT, S(*b*,*a*), makes three distinctive decision at stage n (Ghosh, 1970):

$$\begin{aligned} &\text{Accept } H_o \text{ if } Z_n \le b \\ &\text{Reject } H_o \text{ if } Z_n \ge a \\ &\text{Continue observing data if } b \le Z_n \le a \end{aligned} \tag{13}$$

where the transformed random variable Z_n is the natural logarithm of the probability ratio at stage n:

$$Z_n = \ln \frac{f(E_n \mid H_1)}{f(E_n \mid H_o)} = \ln \frac{f(E_n \mid \sigma_1)}{f(E_n \mid \sigma_o)} \text{ for } n \geq 1 \tag{14}$$

where $f(E_n \mid H_o)$ or $f(E_n \mid \sigma_o)$ is the conditional probabilty of observing the accumuated data set E_n given the assumption that the null hypothese is true. $f(E_n \mid H_1)$ or $f(E_n \mid \sigma_1)$ is defined in a similar fashion. Without any loss of generality, Z_n is defined zero when $f(E_n \mid \theta_1) = f(E_n \mid \theta_o) = 0$. b and a are the two stopping bounds for accepting and rejecting H_o, respectively, and they can be estimated by the following Wald approximations (Wald, 1947):

$$b \cong \ln \frac{\beta}{1-\alpha} \text{ and } a \cong \ln \frac{1-\beta}{\alpha} \tag{15}$$

where α and β the predetermined upper limits for type I and II errors, respectively. When implementing the SPRT, a trade-off must be considered before assigning values for α and β. When there is a large penalty associated with false positive alarms (for example, alarms that shut down traffic over a bridge), it is desirable to keep α smaller than β. On the other hand, for safety critical systems such as nuclear power plants, one might be more willing to tolerate a false positive alarm to have a higher degree of safety assurance. In this case, β is often specified larger than α. Although closed form solutions of a and b are available for several probability models, it has been a standard practice to employ Equation (15) to approximate the stopping bounds in all practical applications. The continuation region $b \leq Z_n \leq a$ is called the *critical inequality* of S(*b*,*a*) at stage n.

If modified observations $\{z_i\}$ $(i = 1, 2, \ldots)$ are defined as follows;

$$z_1 = \ln \frac{f(E_1 \mid \sigma_1)}{f(E_1 \mid \sigma_o)} \text{ and } z_i = \ln \frac{f(E_i \mid \sigma_1) f(E_{i-1} \mid \sigma_o)}{f(E_i \mid \sigma_o) f(E_{i-1} \mid \sigma_1)} \tag{16}$$

then, Z_n becomes:

$$Z_n = \sum_{i=1}^{n} z_i \tag{17}$$

Assuming that E_n has a normal distribution with mean μ and standard deviation σ, z_i can be related to $\varepsilon_y(i)$:

$$z_i = \frac{1}{2}(\sigma_o^{-2} - \sigma_1^{-2})(\varepsilon_y(i) - \mu)^2 - \ln \frac{\sigma_1}{\sigma_o} \tag{18}$$

A normality test such as a norlmality plot on the residaul errors reveals that the distribution of E_n is, in fact, well appromixmated by a Gaussina distribution. Futhemore, the authors also demonstrate that the formulation in Equation (18) is often reliablely applicable to non-Gaussian distributions as well (Sohn et al., 2002).

In a graphical representation of the SPRT S(*b*,*a*), Z_n, which is the cumulative sum of the transformed variable z_i, is continuously plotted against the two stopping bounds *b* and *a*. It should be noted that the mean μ of the distribution is foten assumed to be zero. Even when μ is unkown, the aforementioned procedure is still valid if $\varepsilon_y(i)$ is replaced by x_i:

$$x_i = \left(\sum_{j=1}^{i} \varepsilon_y(j) - i\,\varepsilon_y(i+1) \right) \Big/ \sqrt{i(i+1)} \text{ for } i = 1, 2, \ldots \tag{19}$$

It can be shown that now $\{x_i\}$ has i.i.d. normal distribution with zero mean and the same standard deviation as $\varepsilon_y(i)$.

5. EXPERIMENTAL EXAMPLE

5.1. Description of the Test Structure

An 8-DOF system has been designed and constructed to study the effectiveness of the proposed damage detection procedure. The system is formed with eight translating masses connected by springs. The system employed in this study is shown in Figure 2. Each mass is an aluminum disc of 25.4mm thick and 76.2mm in diameter with a center hole. The hole is lined with a Teflon bushing. There are small steel collars on each end of the discs (Figure 3). The masses all slide on a highly polished steel rod that supports the masses and constrains them to translate only along the rod. The masses are fastened together with coil springs epoxied to the collars that are, in turn, bolted to the masses.

DOFs, springs and masses are numbered from the right end of the system, where the excitation is applied, to the left end as shown in Figure 2. The nominal value of mass 1 (m1) is 559.3 grams. Again, this mass is located at the right end where the shaker is attached. m1 is greater than the others because of the hardware needed to attach the shaker. All the other masses (m2 through m8) are 419.4 grams. The spring constant for all the springs is 56.7 kN/m for the initial condition. Damping in the system is caused primarily by Coulomb friction. Every effort is made to minimize the friction through careful alignment of the masses and springs. A common commercial lubricant is applied between the Teflon bushings and the support rod. Measurements made during damage identification tests were the excitation force applied to m1 and the acceleration responses of all masses. Random excitation was accomplished with a 215N peak force electro-dynamic shaker (Figure 2). The root mean square (RMS) amplitude level of the input was varied from 3 to 7 volts. A Hewlett-Packard 3566A system was employed for data acquisition. A laptop computer was used for data storage and for controlling the data acquisition system. The force transducer used had a nominal sensitivity of 22.48 mv/N, and the accelerometers had a nominal sensitivity of 10 mv/g. The specifications for the data acquisition are summarized in Table 1.

The undamaged configuration of the system is the state for which all springs are identical and have a linear spring constant. Nonlinear damage is defined as an occurrence of impact between two adjacent masses. Damage is simulated by placing a bumper between two adjacent masses so that the movement of one mass is limited relative to the other mass. Figure 3 shows the hardware used to simulate nonlinear damage. When one end of a bumper, which is placed on one mass, hits the other mass, impact occurs. This impact simulates damage caused by the impact from the closing of a crack during vibration. The degree of damage can be controlled by changing the amount of relative motion permitted before contact, and changing the hardness of the bumpers on the impactors. For all damage cases presented, the initial clearance is set to zero.

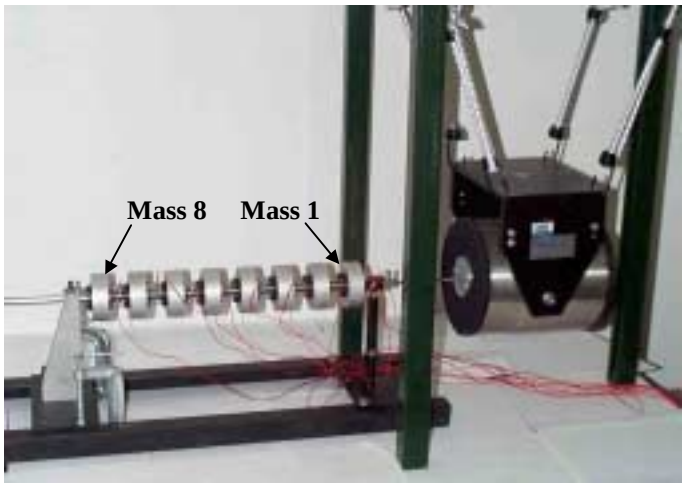

Figure 2: An 8-DOF system attached to a shaker with accelerometers mounted on each mass

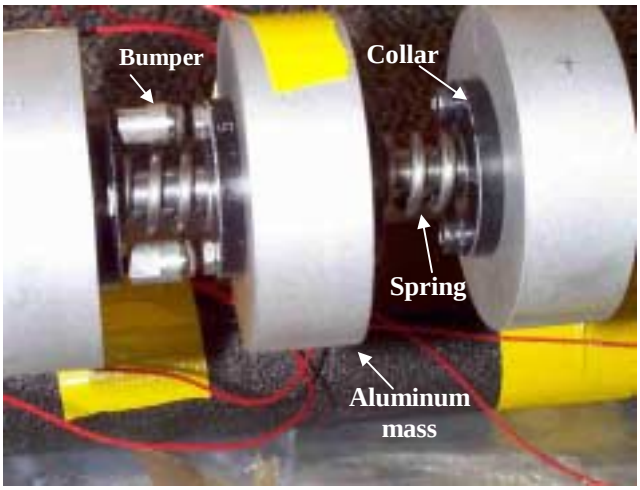

Figure 3: A typical bumper used to simulate nonlinear damage

For the localization study of nonlinear damage, three different damage scenarios are examined varying damage locations and input force levels. This bumper is installed between m1-m2, m5-m6, and m7-m8 for damage cases 1, 2, and 3, respectively. For each damage case, 5 sets of time histories are recorded at an individual input level and the input force varies from 3V to 4, 5, 6, and 7V (except damage case 3, where the input voltage varies from 4V to 7V). Therefore, a total of 25, 25, and 20 time series are recorded for damage cases 1, 2, and 3, respectively. For the undamaged case, 15 sets of time histories are recorded at an individual input level producing a total of 75 time series. Table 2 summarizes the time series studied in this example.

Table 1: Specifications for data acquisition

Time step	0.001953 sec.
Sampling rate	512 Hz
Time period	8 sec.
Frequency resolution	0.125Hz
Number of data points	4096
Filtering	Uniform window
Nyquist frequency	256Hz

Table 2: List of time series employed in this study

Case	Description	Input level	Data # per input	Total data #
0	No bumper	3, 4, 5, 6, 7 Volts	15 sets	75 sets
1	Bumper between m1-m2	3, 4, 5, 6, 7 Volts	5 sets	25 sets
2	Bumper between m5-m6	3, 4, 5, 6, 7 Volts	5 sets	25 sets
3	Bumper between m7-m8	4, 5, 6, 7 Volts	5 sets	20 sets

5.2. Training of the Auto-Associative Neural Network

For this 8-DOF system, the change of excitation levels is the only operation variation. The auto-associative network is first trained for each DOF to reveal the underlying relationship between the AR-ARX coefficients and the unknown excitation levels. It should be noted that an individual neural network is constructed for each DOF. 9 time series out of 15 time series from each input level of the undamaged case are used to train the network. In other words, 45 time series (=9 time series from each input level $\times$ 5 input levels) out of the total 75 time series sets are used for training and the remaining 30 time series are used for validation and testing. From the 45 time series, the coefficients of the AR-ARX model are computed assuming AR(30) and ARX(5,5). That is, the training data set for the network consists of 45 observations with 10 input variables (m=10 and l=45). The input variables are scaled so that each variable has zero mean and unity standard deviation. This scaling weighs all ten variables equally important and is similar to the division of data set by standard deviation often used in the preparation of data for PCA. It should be noted that the excitation level is the main underlying parameter driving the changes of these coefficients. Therefore, the auto-associative neural network with only one node in the bottleneck layer is used to reproduce this training data set (Figure 4).

The auto-associative neural networks with different dimensions in the mapping and de-mapping layers are tried to this training data to determine the best network architecture. In general, the number of nodes in the mapping and de-mapping layers is set to be larger than that of the bottleneck layer ($M_1, M_2 > d$). However, there are no definitive rules for deciding the dimensions of the mapping and de-mapping layers. The complexity of the nonlinear functions, which the neural network represents, primary controls the number of nodes in the mapping and de-mapping layers. If too few nodes are specified in the mapping layers, the accuracy of the neural network might be poor. On the other hand, if too many mapping nodes are provided, the network will be prone to overfitting the stochastic nature of the data rather than learning the underlying functionalities. In practice, the available data might impose constraints on the number of nodes in the hidden layers if the number of training data sets is limited. Otherwise, explicit criteria trading off between the accuracy and the dimension of the hidden layers are often used. Two such criteria are Akaike's Final Prediction Error (FPE) and An Information theoretic Criterion (AIC) (Ljung, 1999):

$$FPE = e(1+N_t/N)/(1-N_t/N) \tag{20}$$

$$AIC = \ln[e] + 2N_t/N \tag{21}$$

where $N_t = (m+d+1)(M_1+M_2)+m+d$ is the total number of weights, $N = lm$ is the number of points in the data, $e = E/(2N)$, and E is the sum of squared errors for all entries in $\mathbf{Y}-\hat{\mathbf{Y}}$. Minimization of these criteria identifies the number of nodes that are neither underparameterized nor overfitted. In this example, a neural network with 3 nodes in each mapping and de-mapping layer has minimized the two criteria on average, and employed for the subsequent SPRT analysis. Several trainings with different initial conditions are required for a given architecture to assure that the global minimum had been achieved. Also, sigmoidal transfer functions were used in all layers except the output layer, which had a pure linear transfer function. The networks employed in this study are conventional feedforward networks and trained by a Levenberg-Marquardt backpropagation with Bayesian regularization (Mackey, 1992) and early stopping (Sarle, 1995). It is reported that the Levenberg-Marquardt algorithm is 10 to 100 times faster than the usual gradient descent method (Hagan and Menhaj, 1994). The Bayesian regularization and early stooping are employed to avoid overfitting the training data set and to improve the generalization of the network.

If the neural network was successfully trained, the output of the bottleneck layer should be closely correlated to the input excitation level because the excitation is the main underlying intrinsic variable causing all the fluctuations. Figure 5 shows a typical relationship between the output of the bottleneck layer and the input level obtained from the network corresponding to m2. Here, the output of the bottleneck layer is an averaged output of 15 times series corresponding to each input levels. The bottleneck output is indeed closely related to the excitation level: the relationship is linear and this is sufficient to reconstruct the input at the output layer. Similar results are also observed from the networks associated with the other measurement points. Therefore, this auto-associative neural network had in a sense revealed the unmeasured operational variation

embedded in this data set. Next, the residual error $\varepsilon_x(t)$ is computed from Equation (3) using the initially estimated AR-ARX coefficient α_i and β_j, and the other residual error $\varepsilon_y(t)$ is obtained from Equation (10) using the neural network prediction $\hat{\alpha}_i$ and $\hat{\beta}_j$.

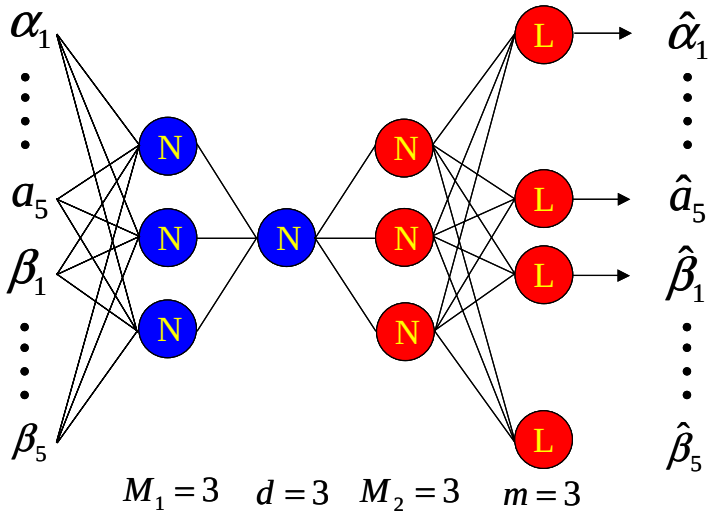

Figure 4: The neural network architecture for the 8-DOF mass-and-spring system

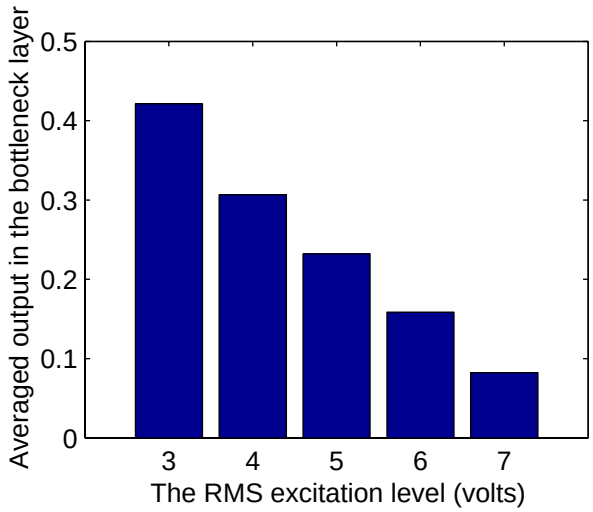

Figure 5: Correlation between the excitation level and the output in the bottleneck layer at mass 2

5.3. Diagnosis Results

Based on the prediction errors, $\varepsilon_x(t)$ and $\varepsilon_y(t)$, computed in the precious step, a simple hypothesis test shown in Equation (11) is performed using the SPRT for an undamaged (case 0) and three damage cases (cases 1-3). For the SPRT analysis, the upper bounds of α and β are set both to 0.01. Furthermore, the two decision boundaries, σ_o and σ_1, are set to $1.3\,\sigma(\varepsilon_x)$ and $1.4\,\sigma(\varepsilon_x)$, respectively. Again, the establishment of these decision boundaries requires the measurement of data sets from a wide range of undamaged cases as well as from a few damage cases. That is, the lower bound should be determined so that most of features extracted from undamaged case do not produce false positive indication of damage, and the upper bound should be selected so that the damage of interest can be detected. Because the sensitive of the extracted

features to damage is structure-dependent, time series signals from a few damage cases might be necessary to choose the upper threshold value.

Table 3 presents the standard deviation ratio, $\sigma(\varepsilon_y)/\sigma(\varepsilon_x)$, for each DOF and all studied cases. The $\sigma(\varepsilon_y)/\sigma(\varepsilon_x)$ values shown in Table 3 are mean values of 75, 25, 25, 20 sample standard deviation ratios for damage cases 0, 1, 2, 3, respectively. If a bumper were introduced at m1, the largest increase in the residual error standard deviation would be expected at the nearest measurement point, m1 However, as shown in Table 3, no significant increase in $\sigma(\varepsilon_y)/\sigma(\varepsilon_x)$ was observed at m1. Instead, the $\sigma(\varepsilon_y)/\sigma(\varepsilon_x)$ value in the next adjacent measurement point, m2, was significantly increased to 1.5322 on average. It is speculated that, because m1 is rigidly connected to the shaker by a rod, the response at this point is masked by the direct influence of the random input. When the bumper was placed at m5, the average $\sigma(\varepsilon_y)/\sigma(\varepsilon_x)$ value in m5 increased to 1.7309, marking the largest increase among all masses (see the fourth row of Table 3). A similar result is observed when the bumper is placed at m7 (see the fifth row of Table 3). A simple chart of the $\sigma(\varepsilon_y)/\sigma(\varepsilon_x)$ values with respect to measurement points seems to reveal the approximate locations of nonlinear damage as well as the existence of the damage.

Next, the SPRT is conducted for all test data, and the diagnosis results are summarized in Table 4. The entries in Table 4 show the rejection number of the null hypothesis $H_0: \sigma(\varepsilon_y) \le 1.3\,\sigma(\varepsilon_x)$ out of all hypothesis tests. For example, when the hypothesis test is conducted at m2 on 75 time series data sets obtained from the undamaged case, the null hypothesis is always accepted for 75 cases (0/75: under "m2" column and "no bumper" row in Table 4). For all damage cases, the number of rejection reaches its peak value near the actual location.

To investigate the effectiveness of the data normalization, a false positive study is conducted by training the auto-associative neural network with only subsets of the 45 time series originally used for training the network. It is observed that the amplification of the input force introduces amplitude-dependent nonlinearity. This input amplification alone might cause a noticeable increase in the standard deviation ratio, $\sigma(\varepsilon_y)/\sigma(\varepsilon_x)$, without the installation of a bumper. The findings of this false positive study are summarized in Table 5. Each entry of Table 5 shows the numbers of false positive indication of damage when the SPRT is applied to times series from the undamaged structure. For example, 16/75 under the column of "m3" and the row of "3V" means that there are 16 false positive indications of damage at m3 out of all tested 75 time series, when only 9 time histories from the excitation level of 3 voltage are used for training the network. The results presented in Table 5 indicate that the success of the proposed data normalization relies on the appropriate inclusion of training data sets from all possible operation and environmental conditions to build the neural network. For comparison, when the network is trained with data from all excitation levels, there is no false positive indication of damage as shown in the second row of Table 4. It should be noted that the measurement of these ambient conditions is often difficult, and they are not necessarily needed for the presented approach. This data normalization employing the auto-associative neural network is also successfully applied to a numerical example of a computer hard disk head, the dynamic properties of which are assumed to be temperature-dependent (Sohn at el., 2001b).

6. CONCLUSIONS

This paper presents a damage identification technique for structural health monitoring, explicitly taking into account changing environmental and operational conditions. A unique integration of the AR-ARX time prediction model, the auto-associative neural network, and the sequential probability ratio test is developed to discriminate the changes of system responses due to ambient operational conditions from those caused by structural damage. The objective of the present damage identification technique is to eschew the physics-based model approaches such as finite element analysis, and therefore pave the way for signal-based techniques applicable to systems of arbitrary complexity. However, the present damage classifier provides an indication only about the presence of damage in a system of interest. This method does not necessarily give information about the location and extent of the damage. That is, the damage classifier only identifies if a new pattern differs from previously obtained patterns in some significant respect. Although the damage assessment problem can be posed with several levels of complexity, the detection of damage presence is arguably the most important step. Once the existence of damage is confirmed, the system can be taken out of service and subjected to detailed inspection to locate and quantify damage. The proposed approach is applied to vibration signals measured from an eight degree-of-freedom mass-and-spring system. Results indicate that the incorporation of the auto-associative network with time series analysis and statistical inference enables one to detect damage even when the system exhibits a range of normal conditions. The development presented here will allow the some progress in in-service monitoring of aerospace, automobile, civil, and mechanical systems, which are subject to various operational and environmental conditions. Such a monitoring system will

be less prone to false-positive indication of damage. To minimize this false indication of damage and develop a more robust monitoring system, the training data set need to be collected over a wide range of environmental and operational conditions of the system. Otherwise, the proposed damage classifier cannot make any definite statement regarding the existence of damage because unusual operational conditions can also have similar effects on the monitoring system. Before the proposed approach could be used with confidence on real structures, several issues need to be addressed. Although the dimension of the bottleneck layer is known *a prior* in this example, this layer size should be also estimated based on model order selection techniques similar to the ones presented in this paper. Often the node numbers of this layer could be initially estimated by grasping the main environmental and operational factors based on observations and engineering judgment. The more principled establishment of the decision boundaries for the sequential probability ratio test also needs to be further investigated considering what degree of damage is statistically significant.

Table 3: The ratio of standard deviations, $\sigma(\varepsilon_y)/\sigma(\varepsilon_x)$, for one undamaged and three damaged cases

Bumper Location	Degree of freedom							
	m1	m2	m3	m4	m5	m6	m7	m8
No Bumper	1.0021	1.0061	1.0185	1.0106	1.0213	1.0278	1.0226	1.0230
Between m1-m2	1.0152	**1.5322**	1.1246	1.0695	1.0461	1.0373	1.0308	1.0287
Between m5-m6	1.0024	1.0116	1.0290	1.0292	**1.7309**	1.2194	1.0510	1.0347
Between m7-m8	1.0018	1.0141	1.0347	1.0186	1.0689	1.1765	**1.7158**	1.3566

*The $\sigma(\varepsilon_y)/\sigma(\varepsilon_x)$ ratios presented are the average values of all input levels. That is, the averages of 75, 25, 25, and 20 individual $\sigma(\varepsilon_y)/\sigma(\varepsilon_x)$ values measured under different input levels are presented.

Table 4: Results of the SPRT analysis for $H_0: \sigma(\varepsilon_y) \le 1.3\,\sigma(\varepsilon_x)$ and $H_1: \sigma(\varepsilon_y) < 1.4\,\sigma(\varepsilon_x)$

Bumper Location	Degree of freedom							
	m1	m2	m3	m4	m5	m6	m7	m8
No Bumper	0/75*	0/75	0/75	0/75	0/75	0/75	0/75	0/75
Between m1-m2	0/25	**25/25**	0/25	0/25	0/25	0/25	0/25	0/25
Between m5-m6	0/25	0/25	0/25	0/25	**23/25**	1/25	0/25	0/25
Between m7-m8	0/20	0/20	0/20	0/20	0/20	2/20	**20/20**	16/20

*Each entry shows the numbers of rejecting the null hypothesis. For example, 0/75 means that the null hypothesis is rejected 0 times out of all tested 75 time series. The SPRT is conducted with α and β = 0.01.

Table 5: Effect of the proposed data normalization on false positive indication of damage

Train data sets	Degree of freedom							
	m1	m2	m3	m4	m5	m6	m7	m8
3V*	0/75**	0/75	16/75	0/75	0/75	28/75	43/75	0/75
4V	0/75	6/75	0/75	0/75	0/75	4/75	0/75	4/75
5V	0/75	0/75	0/75	0/75	0/75	16/75	0/75	0/75
6V	0/75	0/75	0/75	0/75	0/75	0/75	5/75	11/75
7V	0/75	0/75	0/75	0/75	9/75	4/75	0/75	69/75

*The voltage presented denotes the data set used for training the auto-associative neural network. "3V" indicates that 9 time histories obtained only from the RMS excitation level of 3 voltage are used for training.

**Each entry shows the numbers of false positive indication of damage when SPRT is applied to times series from the undamaged case. For example, 0/75 means that there is no false positive indication of damage out of all tested 75 time series.

REFERENCES

1. D. W. Allen, H. Sohn, C. R. Farrar, and K. Worden, "Damage Identification in a Structure Utilizing the Sequential Probability Ratio Test," *SPIE's 7th Annual International Symposium on NDE for Health Monitoring and Diagnostics*, San Diego, CA, USA, March 17-21, 2002.
2. G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, Third Edition, Prentice-Hall, Inc., New Jersey, 1994.
3. G. Cybenko, "Approximation by Superposition of a Signoidal Function," *Math. Control Signal & System*, **2**(4), pp. 303-314, 1989.
4. K. Fukunaga and W. L. G. Koontz, "Application of Karhunen-Loeve Expansion to Feature Selection and Ordering," *IEEE Transactions on Computers*, **C-19** (4), pp. 311-318, 1970.
5. K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, San Diego, CA, 1990.
6. M. T. Hagan, and M. Menhaj, "Training Feedforward Networks with the Marquardt Algorithm," *IEEE Transactions on Neural Networks*, **5**(6), pp. 989-993, 1994.
7. B. K. Ghosh, *Sequential Tests of Statistical Hypotheses*, Addison-Wesley, Menlo Park, CA, 1970.
8. K. C. Gross, and K. E. Humenik, "Sequential Probability Ratio Tests for Nuclear Plant Component Surveillance," *Nuclear Technology*, **93**, pp. 131-137, 1991.
9. K. E. Humenik, and K. C. Gross, "Sequential Probability Ratio Tests for Reactor Signal Validation and Sensor Surveillance Applications," *Nuclear Science and Engineering*, **105**, pp. 383-390, 1990.
10. M. A. Kramer, "Nonlinear Principal Component Analysis Using Autoassociative Neural Networks," *AIChE Journal*, **37**, pp. 233-243, 1991.
11. L. Ljung, *System Identification-Theory for the User*, 2nd Edition, Prentice Hall, Upper Saddle River, New Jersey, 1999.
12. D. C. J. MacKay, Bayesian Interpolation, *Neural Computation*, **4**(3), pp. 415-447, 1992.
13. D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press, Cambridge, MA, 1988.
14. T. D. Sanger, "Optimal Unsupervised Learning in a Single-Layer Linear Feedforward Neural Network," Neural Networks, **2**(6), pp.459-473, 1989.
15. W. S. Sarle, "Stopped Training and Other Remedies for Overfitting", *Proceedings of the 27th Symposium on the Interface*, pp. 1–10, 1995.
16. H. Sohn, M. Dzwonczyk, E. G. Straser, A. S. Kiremidjian, K. H. Law, and T. Meng, "An Experimental Study of Temperature Effects on Modal Parameters of the Alamosa Canyon Bridge," *Earthquake Engineering and Structural Dynamics*, **28**, pp. 879-897,1999.
17. H. Sohn, and C. R. Farrar, "Damage Diagnosis Using Time Series Analysis of Vibration Signals," *Journal of Smart Materials and Structures*, **10**, pp. 446-451, 2001.
18. H. Sohn, C. R. Farrar, N. F. Hunter, and K. Worden, "Structural Health Monitoring Using Statistical Pattern Recognition Techniques," *ASME Journal of Dynamic Systems, Measurement and Control: Special Issue on Identification of Mechanical Systems*, **123**(4), pp. 706-711, 2001a.
19. H. Sohn, C. R. Farrar, and K. Worden, "Novelty Detection Under Changing Environmental Conditions," *SPIE's 8th Annual International Symposium on Smart Structures and Materials*, Newport Beach, CA, USA, March 4-8, 2001b.
20. H. Sohn, D. W. Allen, K. Worden, and C. R. Farrar, Statistical Damage Classification using Sequential Probability Ratio Tests, will be submitted to the *International Journal of Structural Health Monitoring*, 2002.
21. A. Wald, *Sequential Analysis*, John Wiley and Sons, New York, NY, 1947.
22. K. Worden, "Structural Fault Detection Using a Novelty Measure," *Journal of Sound and Vibration*, **201**(1), pp. 85-101, 1997.
23. K. Worden, G. Manson, and N. R. J. Fieller, "Damage Detection Using Outlier Analysis," *Journal of Sound and Vibration*, **229**(3), pp. 647-667, 2000.