# Internal Structure Identification of Random Process Using Principal Component Analysis

Mengqiu Zhang, Rodney A. Kennedy, and Thushara D. Abhayapala
School of Engineering
ANU College of Engineering and Computer Science
The Australian National University
Canberra, ACT, 0200, Australia
email: karan.zhang@cecs.anu.edu.au
rodney.kennedy@anu.edu.au
thushara.abhayapala@anu.edu.au

Wen Zhang
CSIRO
Process Science and Engineering
Lucas Heights
Sydney, NSW, 2234, Australia
email: W.Zhang@csiro.au

***Abstract*—Principal component analysis (PCA) is known to be a powerful linear technique for data set dimensionality reduction. This paper focuses on revealing the essence of PCA to interpret the data, which is to identify the internal structure of the random process from a large experimental data set. We give an explanation of the PCA procedure performed on a generated data set to demonstrate the exact meaning of the dimensionality reduction. Especially, a method is proposed to precisely determine the number of significant principal components for a random process. Then, the internal structure of the random process can be modeled by analyzing the relation between the PCA results and the original data set. This is vital in the efficient random process modeling, which is finally applied to an application in HRTF Modeling.**

## I. Introduction

Advances in data collection and storage capabilities during the past decades have led to an information overload in most sciences, such as engineering, astronomy, biology, remote sensing, economics, and consumer transactions, and so on. Researchers in these areas face new challenges in data analysis not only because of the increase in the number of observations but also the increase in the number of variables associated with each observation. The large complex data set can be difficult to summarize and interpret, which motivates the need for a technique that reduces the effective dimensionality. Principal component analysis (PCA) [1] is such technique with powerful dimensional reduction properties useful in practice.

Smith [2] gave a tutorial in 2002, in which a detailed procedure on how to perform PCA on a data set was elaborated and also an application in computer vision was briefly outlined. Shlens [3] provided an other tutorial in 2009 which focuses on building a solid intuition for how and why principal component analysis works. Abdi and Williams [4] have provided an idea of analyzing the structure of the observations and variables in their upcoming work.

However, these works are largely based on the idea of representing the original dataset, which tends not to be helpful in discovering the exact meaning of the dimensional reduction and the internal structure of the random process under consideration. Moreover, little attempt was made to optimize the method for determining the number of principal components (PCs). The aim of this paper is to answer these questions.

- In Section II we first introduce a new idea of viewing the original dataset as the combination of particular realizations of a random process. This idea is essential to the correct comprehension of the dimensional reduction. For illustration and clearer understanding, we use a synthetic dataset which runs through the whole paper.
- In Section III we give an explanation of the procedure of PCA, in which we clarify the exact meaning of the dimensional reduction. Based on it, we propose a method to precisely determine the number of PCs. Then the internal structure of the random process is identified by discovering the relation between the original dataset and the results of PCA.
- In Section IV we apply the PCA on the CIPIC HRTF Database to determine the appropriate dimension the 3D continuous HRTF model.

## II. Data set: Realizations of Random Process

As we know that the understanding of the properties of a random process usually relies on experiments. Each observed vector obtained by an experiment describes a realization of the random process. Suppose the random process is under consideration and what we have in hand is the experimental data, $N$ realizations of the random process. Then we organize it as $\mathbf{X}$, [1]

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{K1} & x_{K2} & \cdots & x_{KN} \end{bmatrix}, \tag{1}$$

[1]Matrices are denoted in upper case bold, vectors are denoted in lower case bold, and elements are denoted in lower case italic. Matrices, vectors, and elements from the same matrix all use the same letter (e.g., $\mathbf{X}$, $\mathbf{x}$, $x$).

978-1-4244-7907-8/10/$26.00 

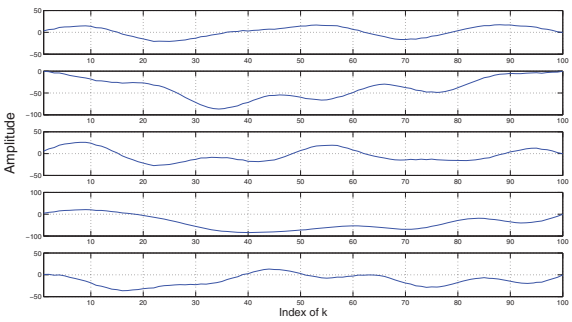


Fig. 1. 5 realizations of the defined random process.

where each column vector is a set of $K$ observable random variables and corresponds to each realization of the random process. Obviously, $\mathbf{X}$ is the aggregation of the experimental data. Doing many experiments to obtain realizations of the random process can be a laborious and time-consuming task. If we can manifest its internal structure based on the finite experimental dataset, representation of any realization of the random process becomes easier. But how could we obtain this structure? The answer is PCA.

For illustration and clearer understanding, we define a random process and generate a dataset. For convenience, we choose 10 orthonormal sinusoidal functions as a basis and generate 10 uncorrelated variables to scale the basis conforming to the Gaussian distribution. Thus a particular realization $\mathbf{x}_n$ of the self-defined random process takes the form of

$$x_n(k) = \sum_{\ell=1}^{10} z_{\ell,n} a_\ell(k) + \bar{x}(k), \tag{2}$$

where $n$ is the index of the realization, $\{\mathbf{a}_\ell\}$ are the orthonormal sinusoidal functions, $z_\ell$ are uncorrelated variables which conform to $\mathscr{N}(0, \sigma_\ell^2)$ ($\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_{10}^2$), $\bar{x}(k)$ is the expected value of $x(k)$ and $k = 1, 2, \cdots, 100$, where $K = 100$. Because $x_n$ is arranged as a column vector, in matrix form, Equation (2) can be written as

$$\mathbf{x}_n = \mathbf{A}\mathbf{z}_n + \bar{\mathbf{x}}, \tag{3}$$

where $\mathbf{x}_n$ is a vector with 100 elements, $\mathbf{A}$ is a $100 \times 10$ matrix, $\mathbf{z}_n$ is a vector with 10 elements, $\bar{\mathbf{x}}$ is a vector with 100 elements, viz.,

$$\begin{bmatrix} x_{1n} \\ x_{2n} \\ \vdots \\ x_{Kn} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1L'} \\ a_{21} & a_{22} & \cdots & a_{2L'} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1} & a_{K2} & \cdots & a_{KL'} \end{bmatrix} \begin{bmatrix} z_{1n} \\ z_{2n} \\ \vdots \\ z_{L'n} \end{bmatrix} + \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_K \end{bmatrix}, \tag{4}$$

where $L' = 10$.

In total, we produce $N$ sets of $\mathbf{z}$. Thus we have $N$ realizations of the random process (here we choose $N = 350$). Five of them are shown in Fig. 1. Therefore we have the original dataset, a $100 \times 350$ matrix, $\mathbf{X}$, which can be written as

$$\mathbf{X} = \mathbf{A}\mathbf{Z} + \bar{\mathbf{X}}, \tag{5}$$

where $\mathbf{X}$ is a $100 \times 350$ matrix, $\mathbf{A}$ is a $100 \times 10$ matrix, $\mathbf{Z}$ is a $10 \times 350$ matrix, $\mathbf{U}$ is $100 \times 350$ matrix with all column vectors the same.

Clearly, the right side of (5) illustrates the internal structure of the random process, where $\mathbf{A}$ is the 10-dimensional sinusoidal basis and $\mathbf{Z}$ are the projections. Once $\mathbf{A}$ and $\mathbf{Z}$ are manifested we can easily represent the process. However in practice what we have is the left side of (5), or equivalently (1). Looking at this data matrix, we can conclude nothing but that at most $K$ variables can be used to represent the random process in a $K$-dimensional space [5]. Fortunately, PCA can be used to extract $\mathbf{A}$ and $\mathbf{Z}$. In next section, we will expose this procedure.

## III. Internal Structure Identification

One approach to a random process is to treat it as a complicate function, which can be expanded to one or several deterministic arguments whose values are random variables with non-deterministic quantities having certain probability distributions. As Karhunen-Loève theorem [7], the expansion basis is deterministic but depends on the process and the coefficients are random variables. Then the internal structure identification procedure turns to find the optimal finite expansion basis and the corresponding random variables.

### A. Subspace Coordinates Identification

In statistics, PCA is a well-known statistical method involving a mathematical procedure that performs an orthogonal linear transformation of the data to a new coordinate system in a lower dimensional space in such a way that the variance of the data in the projected low-dimensional representation is maximized [1], [6]. It is experimentally the optimal transformation for given data in least square terms. In this section we will use PCA to extract the expansion basis, the subspace coordinates, from the experimental dataset.

At first, we need to subtract the mean from each of the data dimensions, which is essential in minimizing the expected value of the mean-squared norm of error [7]. The mean value is calculated by

$$\bar{x}(k) = \frac{1}{N} \sum_{n=1}^{N} x_n(k). \tag{6}$$

Then we have the zero-mean data set $\tilde{\mathbf{X}}$ in which each column is

$$\tilde{\mathbf{x}}_n = \mathbf{x}_n - \bar{\mathbf{x}}_n. \tag{7}$$

After subtracting the mean value from the original dataset, we have the covariance matrix, $\mathbf{C}$,

$$\mathbf{C} = \frac{1}{N-1} \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T, \tag{8}$$

where $\mathbf{C}$ is a $K \times K$ matrix and $\tilde{\mathbf{X}}^T$ is the transposed matrix of $\tilde{\mathbf{X}}$. Now we can compute the eigenvectors $\mathbf{e}_\ell$ and eigenvalues $\lambda_\ell$ for this covariance matrix [8], which satisfies the equation

$$\mathbf{C}\mathbf{e}_\ell = \lambda_\ell \mathbf{e}_\ell, \tag{9}$$

where $\ell = 1, 2, \cdots, K$. The set $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_K\}$ is the set of orthonormal eigenvectors with associated real and non-negative eigenvalues $\{\lambda_1, \lambda_2, \cdots, \lambda_K\}$. These eigenvectors have the characteristics:

- Orthogonality: the covariance between eigenvectors is 0,

$$\text{cov}(\mathbf{e}_\ell, \mathbf{e}_{\ell'}) = 0, \quad \ell \neq \ell', \quad \ell, \ell' = 1, 2, \cdots, K; \quad (10)$$

- Normalization: the variance of the eigenvectors is 1,

$$\text{cov}(\mathbf{e}_\ell, \mathbf{e}_\ell) = 1, \quad \ell = 1, 2, \cdots, K. \quad (11)$$

Once we find the eigenvectors, the next step is to order them by eigenvalues in decreasing order. At this moment, we can say that PCA accomplished the orthogonal linear transformation that transforms the data into a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.

Regarding our generated dataset, the covariance matrix for this dataset is a $100 \times 100$ matrix. Then we have 100 eigenvectors corresponding to 100 eigenvalues. We know that we used the 10-dimensional basis to construct the random process. Obviously, the huge redundancy exists if we introduce 100 eigenvectors to represent the random process. So how many eigenvectors should be chosen as PCs to form the optimal multi-dimensional basis for representing the random process?

A general way of choosing the number of PCs is to set a threshold, where the number $L_1$ is as small as possible while the cumulative energy $P$ is higher than a certain threshold $\xi$,

$$P = \frac{\sum_{\ell=1}^{L_1} \lambda_\ell}{\sum_{\ell=1}^{L} \lambda_\ell} \geq \xi, \quad (12)$$

where $(\lambda_1 \geq \cdots \geq \lambda_{L_1} \geq \cdots \geq \lambda_L)$. Then the first $L_1$ eigenvectors are chosen as PCs, $\mathbf{e}_\ell$, where $\ell = 1, \cdots, L_1$. The chosen PCs form a $L_1$-dimensional subspace basis for representing realizations of the random process. Apparently, $K - L_1$ eigenvectors of lesser significance are ignored because they carry small information about the process. But unfortunately, in practice, it is actually not the case.

Recall the example given in Section II, we have 100 eigenvectors for the generated dataset. Then we set $\xi = 0.98$, which means that only 2 percent of information are discarded. And the first 9 eigenvectors are chosen as the PCs ($L_1 = 9$). The 9 PCs do satisfy most realizations of the random process, but may fail to represent a few of realizations. Fig. 2 exhibits the reconstructed errors, the expression is given later in (21), of all realizations of the defined random process. We can see that most reconstructed errors are acceptable while few particular realizations are very large, even more than 40%. In other wards, the 9-dimensional subspace is not a complete space [9] for reconstructing the random process. It is easy to understand because the dataset is generated by 10-dimensional basis. Thereupon, the method of setting threshold to determine the number of PCs might satisfy some cursory applications.

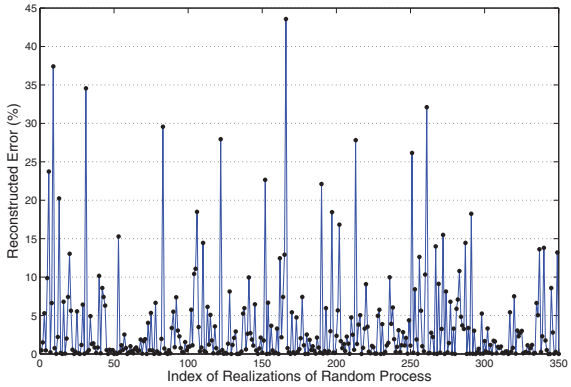


Fig. 2. Reconstructed Errors for all realizations of the random process when $\xi$ = 0.98.

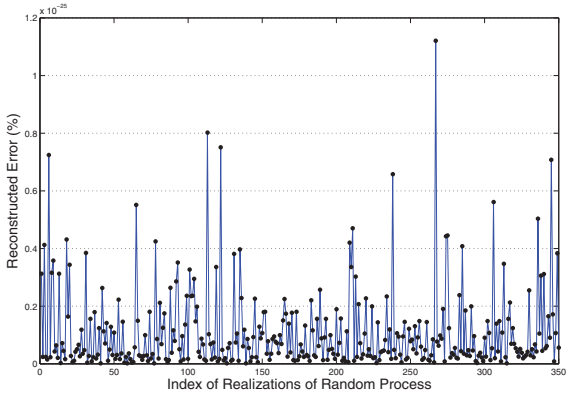


Fig. 3. Reconstructed errors for all realizations of the random process when a complete subspace basis is used to realize a complicate random process.

But it is undoubtedly weak in exposing the real structure of the random process.

To achieve the goal of disclosing the random process structure, we need to precisely determine the dimensionality of the subspace basis. After performing PCA on the generated dataset, we find that if we add one more eigenvector (the 10th eigenvector) into the 9-dimensional subspace the random process can be fully expressed. The reconstructed error of any realization of the random process will be extremely small as shown in Fig. 3. And it would not decrease significantly if adding more eigenvectors into the 10-dimensional subspace. Then a conclusion can be drawn that the 10-dimensional subspace is the complete space for reconstructing the defined random process. It is reasonable because its dimensionality is exactly the same as that of the basis using in the original dataset generation.

**Remarks**: The conclusion is that the overall reconstructed error is the crucial factor to judge whether the chosen PCs form a complete subspace or not. Thus we propose to set a threshold at first to find $L_1$-dimensional subspace and then enlarge it till the overall reconstructed error does not decrease

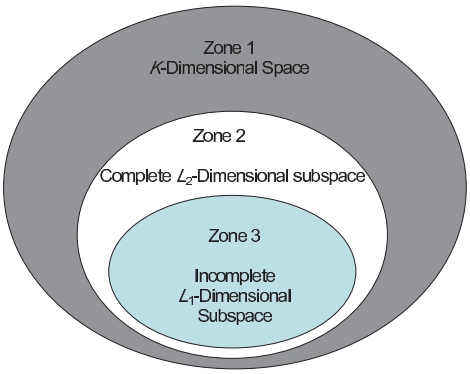


Fig. 4. Notion of Dimensionality Reduction.

in orders (for example from $10^{-5}$ to $10^{-10}$). Assuming the $L_1$-dimensional subspace is enlarged to the $L_2$-dimensional space ($L_1 = 9$ and $L_2 = 10$ in our example). Then we have a $K \times L_2$ PC matrix,

$$\mathbf{E} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1L_2} \\ e_{21} & e_{22} & \cdots & e_{2L_2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{K1} & e_{K2} & \cdots & e_{KL_2} \end{bmatrix}, \tag{13}$$

in which each column is the principal component, the new sub-space coordinates.

Definitely, $L_1 < L_2 < K$, whose relation is depicted in Fig. 4. The notion of dimensionality reduction can be seen clearly in this figure. Zone 1 denotes all eigenvectors of the covariance matrix, which is the $K$-dimensional space. Certainly we can use this $K$-dimensional basis to represent the random process without losing any information. But the representation is not efficient because of the existence of the redundancy. Zone 3 denotes the incomplete $L_1$-dimensional subspace determined by setting a threshold. But the representation is not fully credible because of the missing information. Zone 2 denotes the complete $L_2$-dimensional subspace (the optimum basis we sought) by which we can represent the random process completely and efficiently as well.

### *B. Random Variables Determination*

The $L_2$ orthonormal PCs are the new sub-space basis vectors sought by PCA on which the random process is projected. The projections are the random variables in Karhunen-Loève expansion, which is given by

$$w_{\ell,n} = \sum_{k=1}^{K} \tilde{x}_n(k) e_\ell^*(k), \tag{14}$$

where $*$ represents complex conjugate, $\ell = 1, 2, \cdots, L_2$, and $n = 1, 2, \cdots, N$. For a particular realization of the random process, $L_2$ weights are needed to scale $L_2$ PCs. Then for $N$ realizations, we have a $L_2 \times N$ weights matrix,

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & \cdots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{L_2 1} & w_{L_2 2} & \cdots & w_{L_2 N} \end{bmatrix}, \tag{15}$$

in which each column is a set of the zero-mean uncorrelated variables and each row is a set of realizations for a particular variables. The weights of first row correspond to the first PC, the weights of second row correspond to the second PC, and so on. Generally the variables show the following distribution properties,

- Zero-mean: the mean value of each row of $\mathbf{W}$ is 0,

$$E(\mathbf{w}_\ell) = 0; \tag{16}$$

- Uncorrelated: the covariance between different rows of $\mathbf{W}$ is 0,

$$\text{cov}(\mathbf{w}_\ell, \mathbf{w}_{\ell'}) = 0, \quad \ell, \ell' = 1, 2, \cdots, L_2, \quad \ell \neq \ell'; \tag{17}$$

- Variance: the variances of the first row of $\mathbf{E}$, which corresponds to the first PC, is largest and it decreases along with the increase of the order of the row.

For the previous example, 10 random variables are calculated for each realization. Then we have a $10 \times 350$ variables matrix.

### *C. Internal Structure of Random Process*

Once we have the PCs and the weights, the particular realization of the random process can be given by

$$\hat{x}_n(k) = \sum_{\ell=1}^{L_2} w_{\ell,n} e_\ell(k) + \bar{x}_n(k). \tag{18}$$

We can rewrite (18) as

$$\hat{\mathbf{x}}_n = \mathbf{E}\mathbf{w}_\mathbf{n} + \bar{\mathbf{x}}, \tag{19}$$

where

- $\hat{\mathbf{x}}_n$, a vector with $K$ elements, is the particular realization of the random process;
- $\mathbf{E}$ is the $K \times L_2$ PC matrix as shown in (13);
- $\mathbf{w}_n$ is the weights vector with $L_2$ elements, as the column vector shown in (15);
- $\bar{\mathbf{x}}$ is the mean value vector of $\hat{\mathbf{X}}$ or $\mathbf{X}$ with $K$ elements, $[\bar{x_1}, \bar{x}_2, \cdots, \bar{x}_K]^T$.

Equation (18) demonstrates that the random process is projected to the new $L_2$-dimensional sub-space coordinates (PCs) and scaled by $L_2$ uncorrelated weights. In other words, we can use $L_2$ uncorrelated random variables to represent the random process. Indeed, this is the central idea behind dimensional reduction.

Then for all realizations of the random process, we have

$$\hat{\mathbf{X}} = \mathbf{E}\mathbf{W} + \bar{\mathbf{X}}, \tag{20}$$

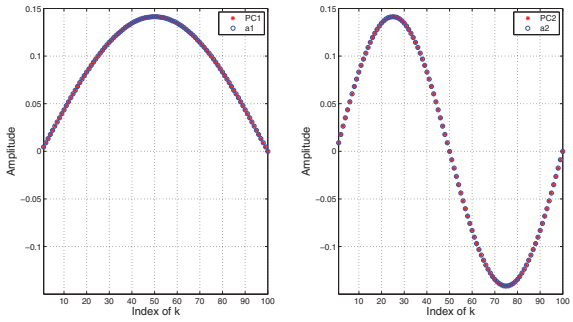


Fig. 5. Comparison of $\mathbf{E}$ and $\mathbf{A}$: the left panel shows the first PC and the first column vector of $\mathbf{A}$; the right panel shows the second PC and second column vector of $\mathbf{A}$.

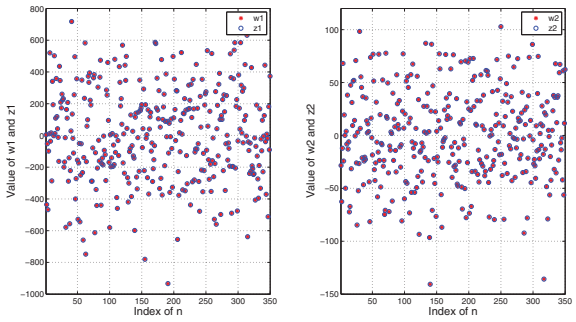


Fig. 6. Comparison of $\mathbf{w}$ and $\mathbf{z}$: the left panel shows the weights for the first PC of all realizations and all values of the first variable; the right panel shows the weights for the second PC of all realizations and all values of the second variable.

where $\hat{\mathbf{X}}$ is a $K \times N$ matrix, $\mathbf{E}$ is a $K \times L_2$ matrix, $\mathbf{W}$ is a $L_2 \times N$ matrix, $\bar{\mathbf{X}}$ is $K \times N$ matrix with all column vectors the same as $\bar{\mathbf{x}}$.

Next we shall examine the new transformation by the mean-squared norm of error (we call it the reconstructed error) given by

$$err_n = \frac{\left\|\mathbf{x}_n - \hat{\mathbf{x}}_n\right\|^2}{\left\|\mathbf{x}_n\right\|^2}, \tag{21}$$

where $n = 1, 2, \cdots, N$. As discussed in Section III-A, the overall reconstructed error will be extremely small if the complete sub-space basis is used to represent the random process.

We notice that the new transformation (20) has the similar form to (3). Because the original dataset is generated by ourselves we know exactly what $\mathbf{A}$, $\mathbf{z}$ and $\mathbf{u}$ are. So it is convenience to compare them. The results are listed as follow.

- $\mathbf{E}$ and $\mathbf{A}$ are both $100 \times 10$ matrix. And they are identical. We only present the first two PCs and the first two column vectors of $\mathbf{A}$ in Fig. 5. This fact illuminates that the PCs sought by PCA are the functions what we use in the random process definition.
- $\mathbf{z}_n$ and $\mathbf{w}_n$ are both vectors with 10 elements. They are also identical. Fig. 6 presents the weights for the first two PCs of all realizations and all values of the first two variables of $\mathbf{z}$. This illuminates that the weights are the random variables what we use to scale the function basis in the random process definition.
- $\bar{\mathbf{x}}$ equals to $\mathbf{u}$, the expected value of $\mathbf{X}$.

From the above analysis, we know that PCA can extract the sub-space basis hiding in the experimental dataset. The manner of the PCA results is akin to visualize an ellipsoid in three-dimensional Cartesian coordinate system where the three axes can be treated as PCs and the three coordinates as the random variables, which denote the size, the shape and the position of the ellipsoid. Now the internal structure of the random process is unfolded before us: the random process is projected on the optimal sub-space coordinates and specialized by the random variables.

## IV. Application

Looking back, before PCA, what we have is $N$ column vectors of $\mathbf{x}$, the original dataset $\mathbf{X}$, while after PCA, what we have are PCs $\mathbf{E}$, the orthonormal sub-space coordinates, sets of weights $\mathbf{w}_n$, the random variables, and the expect value vector $\bar{\mathbf{x}}$. Therefore, the understanding of the properties of the new orthonormal coordinates and the distribution of the random variables can help us to establish a mathematical model to represent the random process.

In this section, we apply the above technique to determine the appropriate dimensionality of the far-field 3D HRTF (Head-Related Transfer Function) model [10], in which the HRTF spatial components are expanded using spherical harmonics and the frequency components using Fourier Spherical Bessel (FSB) series.

According to [10], the spatial decomposition of HRTF can be written as

$$\hat{H}(\theta_s, \phi_s, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_n^m(k) Y_n^m(\theta_s, \phi_s), \tag{22}$$

where $\theta_s$ and $\phi_s$ define the sound source position, $k = 2\pi f/c$ is the wavenumber, $Y_n^m(\theta_s, \phi_s)$ are the spherical harmonics characterized by degree $n$ and order $m$, $\alpha_n^m(k)$ are the spherical harmonic coefficients of the equivalent source field at wavenumber $k$. Further, the frequency decomposition is applied on $\alpha_n^m(k)$, which is

$$\alpha_n^m(k) = \sum_{q=1}^{Q} A_{n,q}^m j_n\left(\frac{Z_q^{(n)}}{k_{max}} k\right), \tag{23}$$

where $A_{n,q}^m$ are the FSB coefficients, $j_n(\cdot)$ is the spherical Bessel function of the degree $n$, $Z_q^{(n)}$ are the positive roots of $j_n(\cdot)$. Then we examine how to determine the spatial dimensionality $N$ and frequency dimensionality $Q$. In [10], the spatial dimensionality $N$ varies with the wavenumber $k$ and the radius of the equivalent sound source field while the frequency dimensionality $Q$ is set to a fixed number. But the result of PCA indicates that fewer number of spherical

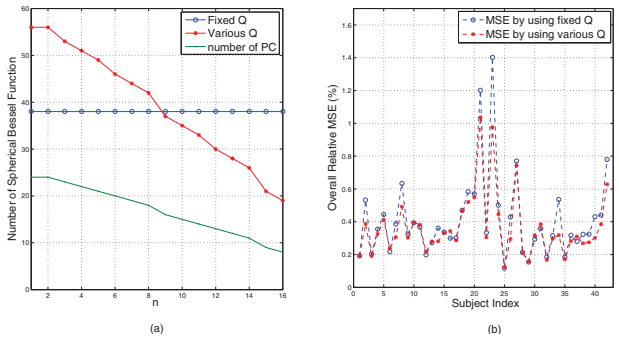


Fig. 7. Comparison of two methods to reconstruct the spherical harmonic coefficients. (a) The number of spherical Bessel functions used to reconstruct $\alpha_{n,i}^m(k)$ and the number of PC. (b) The overall relative mean square error of $\hat{\alpha}_{n,i}^m(k)$ for all subjects.

Bessel functions could be used to represent spherical harmonic coefficients at higher degree .

We use CIPIC HRTF Database [11] here to do the analysis. The procedure is summarized as follow. At first we calculate $\alpha_{n,i}^m(k)$ ($i$ is the subject index) for all 42 subjects. Then PCA is performed on $\alpha_{n,i}^m(k)$ along $k$ for a particular $n$ and $m$. We find that the number of PC changes with $n$ no matter what $m$ is. The larger value of $n$, the smaller number of PC, which suggest that we could use a greater number of spherical Bessel functions to reconstruct spherical harmonic coefficients at low degree of $n$ and less for those at high degree of $n$. However, PCs extracted from PCA procedure are orthonormal while spherical Bessel functions are not, which means more spherical Bessel functions are needed than PCs. Fig. 7 shows the comparison of two methods to reconstruct the spherical harmonic coefficients. The blue circles in the left panel indicate that the equal number of $Q$ is used to reconstruct $\alpha_{n,i}^m(k)$ of any $n$. The reconstructed results, which is evaluated by the overall relative mean square error (MSE),

$$\varepsilon(i) = \frac{\sum_{n=0}^{N}\sum_{m=-n}^{n}\sum_{k=1}^{K}|\alpha_{n,i}^m(k) - \hat{\alpha}_{n,i}^m(k)|^2}{\sum_{n=0}^{N}\sum_{m=-n}^{n}\sum_{k=1}^{K}|\alpha_{n,i}^m(k)|^2}, \quad (24)$$

are shown in the right panel by the blue circles. The red stars in the left panel display that the different number of spherical Bessel functions, around 2.3 times the number of PC shown by the green dots, are used to reconstruct $\alpha_{n,i}^m(k)$. Then the total number of FSB coefficients reduced by about $15\%$ while with a better performance as shown by the red stars in the right panel.

## V. Conclusion

In this paper we introduce an idea of viewing the original dataset as the combination of particular realizations of the random process. Then we elaborated the PCA procedure by laying great emphasis on exploring the real meaning of the dimensional reduction which leads us clarify the internal structure of the random process. The clarification of the internal structure is important in the random process understanding and then modeling.

## References


[1] I. T. Jolliffe. *Principal Component Analysis*. Springer series in statistics. Springer verlag, 175 Fifth Avenue, New York, NY 10010, USA, 2nd edition, 2002.

[2] L. I. Smith. A tutorial on principal components analysis. *Cornell University, USA*, 51:52, 2002.

[3] J. Shlens. A tutorial on principal component analysis. *Systems Neurobiology Laboratory, University of California at San Diego*, 2009.

[4] H. Abdi and L. J. Williams. Principal component analysis. *Computational Statistics*, 2, 2010. in press.

[5] G. Strang. *Introduction to linear algebra*. Wellesley-Cambridge, Wellesly, MA, USA, 3rd edition, 2003.

[6] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417–441, 1933.

[7] L. Franks. *Signal Theory*. Prentice-Hall, Englewood Cliffs, N.J., 1969.

[8] H. Anton and C. Rorres. *Elementary Linear Algebra Applications Version*. Wiley India Pvt. Ltd., New York, 6th edition, 2008.

[9] K. Atkinson and W. Han. *Theoretical numerical analysis: A functional analysis framework*. Springer, London, 3rd edition, 2009.

[10] W. Zhang, R. A. Kennedy, and T. D. Abhayapala. Efficient continuous hrtf model using data independent basis functions: Experimentally guided approach. *IEEE Trans. Audio, Speech and Language Processing*, 17:819–829, 2009.

[11] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The CIPIC HRTF database. In *Proc.IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, pages 99 – 102, New Paltz, NY, USA, October 2001.