


# Optimized phase-space reconstruction for accurate musical-instrument signal classification

Yina Guo<sup>1</sup>  · Qijia Liu<sup>1</sup> · Anhong Wang<sup>1</sup> · Chaoli Sun<sup>1</sup> ·  
Wenyan Tian<sup>1</sup> · Ganesh R. Naik<sup>2</sup> · Ajith Abraham<sup>3</sup>

Received: 4 May 2016 / Revised: 14 September 2016 / Accepted: 27 September 2016 /  
Published online: 7 October 2016  
© Springer Science+Business Media New York 2016

**Abstract** Traditional musical-instrument classification methods mainly use regions in the time or/and frequency characteristics, cepstrum characteristics, and MPEG-7 characteristics, and they often lead to erroneous classification. Therefore, there is need to develop a more suitable method that is more applicable to the nonlinear characteristics of musical-instrument signals and can avoid the abovementioned problems. In this paper, a musical-instrument classification method that couples the optimized phase-space reconstruction (OPSR) with a flexible neural tree (FNT) is proposed. As per nonlinear dynamic theory, a principal component analysis and correlation coefficient are used to optimize the phase-space reconstruction (PSR) method. Multidimensional PSR results for different musical-instrument signals are extracted as the main components, and the dimensionality is reduced by the OPSR method. A probability density function (PDF) is introduced in the feature extraction step to differentiate musical instruments according to the phase-space-reconstructible characteristics. A FNT is adopted as a classifier to tackle the variability in musical-instrument signals and to improve the adaptive ability of various target classification problems. Experimental testing has been conducted to show that the proposed OPSR–PDF–FNT algorithm gives superior performance over other comparable algorithms and can classify 12 musical instruments with an accuracy of 98.2 %.

**Keywords** Musical-instrument classification · Phase-space reconstruction · Principal component analysis · Flexible neural tree

---

✉ Yina Guo  
zulibest@163.com

<sup>1</sup> Taiyuan University of Science and Technology, Taiyuan 030024 Shanxi, China

<sup>2</sup> Centre for Health Technologies (CHT), University of Technology, Sydney 2007, Australia

<sup>3</sup> Machine Intelligence Research Labs (MIR Labs), Scientific Network for Innovation and Research Excellence, P.O. Box 2259, Auburn, WA 98071-2259, USA

## 1 Introduction

Music is broadly divided into vocal and instrumental music. Music classification is currently a popular topic and is of great importance in life. McKay proposed the automatic music classification software jMIR for music information retrieval (MIR) research [21]. jMIR gives a good performance for the music classification of melodies, pitches, and rhythmic patterns. However, with regard to musical-instrument classification, jMIR is mainly focused on the identification of particular instruments. Thus, it is necessary to explore a more suitable instrument classification method.

In general, musical-instrument families comprise the wind, string, and keyboard families. The timbre or tone quality is used to distinguish different musical instruments. However, the tone is a subjective attribute that cannot be accurately described using certain physical parameters. Methods based on the time or/and frequency, the cepstrum, higher-order spectra, and MPEG-7 are commonly used in feature extraction for classifying musical instruments [1, 5, 10, 23]. Several conventional machine learning methods such as the k-nearest neighbor algorithm, a support vector machine (SVM), an artificial neural network, a Gaussian mixture model, and a hidden Markov model (HMM) [1, 5, 23] are employed for musical-instrument classification.

Musical-instrument signals typically exhibit nonlinear characteristics. Eronen et al. applied mel-frequency and linear prediction cepstral coefficients and incorporated them in a k-nearest neighbor algorithm to classify 16 different western orchestral instruments [10]. The best performance of the method was a solo tone recognition rate of 35 % for individual instruments and 77 % for families. Agostini et al. used the spectral characteristics of sound with SVMs and quadratic discriminant analysis to classify 27 musical instruments [1]. The experiment results show successful classification rates close to 70 % for family instruments. Bhalke et al. proposed the classification and recognition of monophonic isolated musical-instrument sounds using higher-order spectra such as a bispectrum and trispectrum [5]. Their experimental results show that higher-order spectrum-based variable features improve the recognition accuracy to 88.59 % for family instruments and to 71.93 % for individual instruments. Patil et al. adopted biomimetic spectrotemporal features and an SVM to identify of musical instruments [23]. The musical-instrument recognition accuracy is up to 97 % for isolated notes and solo phrases. However, a single classifier of an artificial neural network (ANN), SVM, or HMM may not provide a high recognition performance for family instruments. In order to improve the performance of a single classifier, multiclassifier fusion methods have been widely used. Benetos et al. presented an automatic musical-instrument identification method to recognize 20 instruments using MPEG-7 descriptors and a variety of classifiers [4]. Their method has a high classification accuracy in the range of 88.7–95.3 %. However, these feature-extraction methods are not sufficient for describing the nonlinear characteristics of musical instruments and often result in the erroneous classification of different family instrument members, such as the misclassification of clarinet, oboe, and cello signals [24].

Phase-space reconstruction (PSR) is a delay embedding theorem used in nonlinear dynamic system research and was proposed by Takens in 1981 [25]. In mathematics, the delay embedding theorem provides the conditions under which a chaotic dynamical system can be reconstructed from a sequence of observations of the state of a dynamical system [26]. Now, it is widely employed in medicine [20, 27], astronomy [15], electricity [29], and water disposal [16]. On the basis of PSR and the multiclassifier fusion method, Rui et al. discussed the PSR of audio time sequences produced by different instruments [24]. Their experimental results

indicate that the recognition accuracy of their method is 97.8 % for instrument families and 90.4 % for individual instruments. These methods can demonstrate the nonlinear characteristics of musical instruments; however, the large amount of phase-space-reconstructed data is needed to achieve a good reduction in the dimension. Furthermore, the multiclassifier fusion method makes the musical-instrument classification process complex, although it improves the recognition accuracy.

A flexible neural tree (FNT) was first proposed by Chen [7]. A FNT is a fuzzy model that was initially proposed to solve the highly structure-dependent problems of an artificial neural network [7, 8]. The FNT model is computed as an irregular flexible multilayered feed-forward neural network. On the basis of predefined instruction/operator sets, a FNT can be created and evolved [8]. As an algorithm of computational science [6–8], a FNT outperforms other classifiers including an SVM and ANN [7] and is mainly applied in computer modeling and pattern recognition for medicine [6, 30], biosignal processing [12, 14], and image processing [28].

In this paper, a musical-instrument classification method that combines an optimized PSR (OPSR) with a FNT is proposed. As per nonlinear dynamic theory, a principal component analysis (PCA) and correlation coefficient are used to optimize the PSR method. Multidimensional PSR results of different musical-instrument signals are extracted as the main components, and the dimensionality is reduced using OPSR. Thus, the large amount of phase-space-reconstructed data can be reduced in dimension. A probability density function (PDF) is introduced in the feature extraction step to differentiate the musical instruments according to the phase-space-reconstructible characteristics.

A FNT is adopted as the classifier to tackle the variability in the musical-instrument signals. The flexible tree structure makes it possible to automatically choose input variables and reduce the dimension of the input space. From this, the FNT can instead of the combination of classifiers, and can reduce the complexity of musical instrument classification process. The adaptive ability of various target classification problems is also improved. The classification performance of the OPSR–PDF–FNT algorithm is compared with other comparable algorithms through experiments; the experimental results show that the proposed algorithm outperforms the other algorithms with a higher recognition rate (Accuracy) and a lower root mean squared error (RMSE).

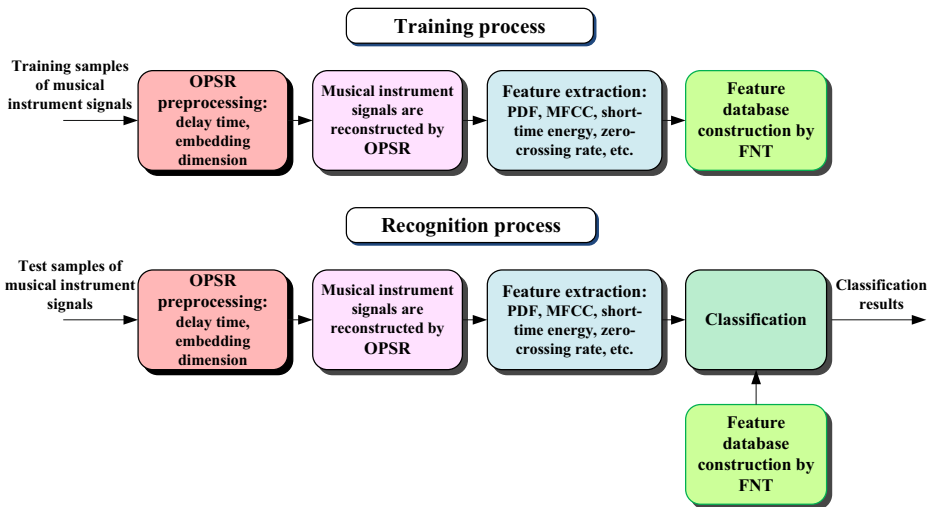
## 2 Methodology

### 2.1 Algorithm

Figure 1 shows the training and recognition processes of musical-instrument signal classification. The training process consists of OPSR preprocessing, OPSR, feature extraction, and feature database construction using a FNT. The recognition process consists of OPSR preprocessing, OPSR, feature extraction, and classification with the feature database. The RMSE and classification rate are used to evaluate the performance of the OPSR–PDF–FNT algorithm for musical-instrument classification [14, 18, 24].

#### 2.1.1 OPSR preprocessing

A one-dimensional time series  $x = (x_1, x_2, \dots, x_K)^T$  that demonstrates nonlinear characteristics is considered. In order to extract useful information from a time series,  $x$  is



**Fig. 1** Flowchart of the musical-instrument signal classification process

expanded to an  $m$ -dimensional embedded space, and the reconstructed phase-space vector  $Y$  is

$$Y = (y_1, y_2, \dots, y_N) = \begin{pmatrix} x(1) & x(2) & \dots & x(N) \\ x(1 + \tau) & x(2 + \tau) & \dots & x(N + \tau) \\ \vdots & \vdots & \ddots & \vdots \\ x(1 + (m-1)\tau) & x(2 + (m-1)\tau) & \dots & x(N + (m-1)\tau) \end{pmatrix} \quad (1)$$

where  $\tau$ ,  $m$ , and  $K$  are the delay time, the embedding dimension, and the length of  $x$ , respectively. The number of phase vectors in the phase space is  $N = K - (m - 1)\tau$ . For reconstructing  $x$  in a higher-dimensional space,  $\tau$  and  $m$  are critical parameters in PSR [24].

**Delay time  $\tau$**  The dynamic differences between the respective components in  $Y$  can be obtained from the evolution of the signal source in  $\tau$ . The appropriate selection of  $\tau$  can make the coordinates neither linear nor fully independent, and it plays a vital role in data model performance. An appropriate value of  $\tau$  is derived using the average mutual information method [20, 27]:

$$I(\tau) = \sum_{n=1}^N P(x_n, x_{n+\tau}) \log \frac{P(x_n, x_{n+\tau})}{P(x_n)P(x_{n+\tau})} \quad (2)$$

where  $P(x_n)$  and  $P(x_n, x_{n+\tau})$  are probabilities. The optimum  $\tau$  is the delay time corresponding to the first local minimum value of  $I(\tau)$ .

**Embedding dimension  $m$**  The aim of selecting the embedding dimension is to make the original dynamic system and the reconstructed phase space topology equivalent. An appropriate embedding dimension can describe the characteristics of the original dynamic system accurately and reduce the influence of the calculation and noise.  $m$  is derived by using the false nearest neighbor (FNN) method [15, 16, 29].

In an  $m$ -dimensional phase space, each phase point  $X(i) = [x(i), x(i + \tau), \dots, x(i + (m - 1)\tau)]$  has a closest point  $X^{NN}(i)$  within a certain distance, and the distance between these two points is given by

$$R_m(i) = \|X(i) - X^{NN}(i)\| \tag{3}$$

When the embedding dimension of the phase space is between  $m$  and  $m + 1$ , the distance of the two points will change. In this case, the distance between the points is  $R_{m+1}(i)$ , which is given as

$$R_{m+1}^2(i) = R_m^2(i) + \|X(i + \tau m) - X^{NN}(i + \tau m)\|^2 \tag{4}$$

When  $a(i, m) = \|X(i + \tau m) - \frac{X^{NN}(i + \tau m)}{R_m(i) > R_\tau}\|$ ,  $X^{NN}(i)$  is the FNN point of  $X(i)$ , where  $R_\tau$  is a threshold that lies between 10 and 50. The proportion of FNN points can be calculated by identifying the falsity of the neighboring points of each phase vector in the  $m$ -dimensional space. When the proportion of FNN points is less than 5 %, the phase-space trajectory is fully open, and  $m$  is the optimum embedding dimension.

### 2.1.2 OPSR

OPSR can improve the dimension reduction performance of  $Y$ . A principal component analysis (PCA) is applied to reduce the dimensionality of  $Y$ .  $Y$  is projected into the reduced space defined by only the first  $L$  singular vectors  $W_L$ :

$$A = W_L^T Y = \sum_L V^T \tag{4}$$

where  $\sum_L = I_L \times m \sum$ , and  $I_L \times m$  is an  $L \times m$  rectangular identity matrix. The matrix  $\sum$  is an  $m \times N$  rectangular diagonal matrix with nonnegative real numbers on the diagonal, and the  $N \times N$  matrix  $V$  is the matrix of the eigenvectors of  $Y^T Y$  [11, 18].

After the first dimension reduction by PCA, the components of the dimension-decreasing vector  $A$  are mainly independent. However, the amount of data in  $A$  is still large and increases the complexity of the subsequent experiment. In order to reduce the dimensionality of  $A$  and shorten the running time of the program, the correlation coefficients between the components of  $A$  and the original musical-instrument signal  $x$  are adopted. The correlation coefficient can be expressed as [13]

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(c_i(j) - \bar{c}(j))}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (c_i(j) - \bar{c}(j))^2}} \tag{5}$$

When the correlation coefficient is less than the correlation coefficient criterion, the component of  $A$  will be discarded. The correlation coefficient criterion is formulated in Section 2.2. After the second dimension reduction by the correlation coefficient,  $A$  is changed to an  $M \times N$  rectangular matrix  $C$ .

### 2.1.3 Feature extraction

For musical-instrument classification, the extraction of features that reflect the essential characteristics of the musical-instrument signals is important. For describing the differences in various musical instruments more accurately, except for some traditional acoustic parameters, the mel-frequency cepstral coefficient (MFCC), short-time energy, zero crossing rate, and linear prediction coefficient are used to extract the characteristics of musical-instrument signals. In addition, a PDF is used to depict the intensity distribution of each trajectory in the phase space. The extracted features of the musical-instrument signals are the inputs of the musical-instrument recognition model. The expression of the PDF is as follows:

$$p(c) = \frac{1}{N} \sum_{i=1}^N \frac{1}{h_M} k\left(\frac{c-c_i}{h}\right) \tag{6}$$

where  $x$  is the center, and  $h$  is the side length of  $G$ .  $G$  is a tiny cube.  $k(u)$  is a kernel function.  $N$  is the number of samples, and  $M$  is the number of data dimensions. According to each component of  $C$ , we can sample and compute the PDF. From this basis, the feature set  $H$  with  $M$  lines and  $B$  columns can be obtained by PDFs. Note that,  $B < N$  [1, 5].

### 2.1.4 Feature database construction using a FNT

The musical-instrument recognition model is approximated using a neural tree model with predefined instruction sets. The instructions of the root node, hidden nodes, and input nodes are selected from three instruction sets [7, 8]. We have used two instructions in the experiments. The instruction sets are as follows (see Fig. 2):

$+_i (i=2, 3, \dots, N)$  denotes the instructions of the nonleaf nodes that take  $i$  arguments, and  $h_1, h_2, \dots, h_n$  represent the instructions of the leaf nodes that take no other arguments.

Figure 3 clearly shows that the output of  $+_n$ , which is also called a flexible neuron operator, is calculated as a flexible neuron model with  $n$  arguments.

**Fig. 2** Flexible neural tree with function instruction sets  $I = \{+_2, +_3, +_4, +_5, h_1, h_2, h_3\}$

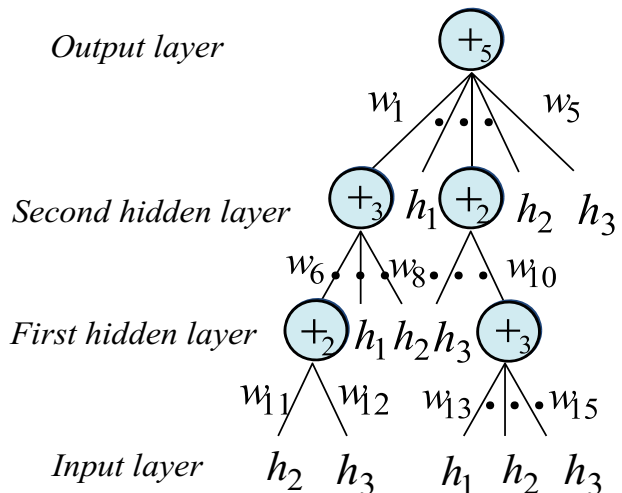
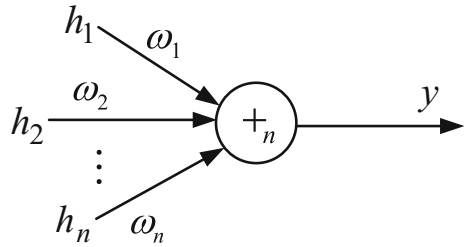


Fig. 3 Flexible neuron operator



When a nonleaf instruction  $+_i (i=2, 3, \dots, N)$  is selected,  $i$  real values are evolved automatically and used to demonstrate the connection strength between  $+_i$  and its children. The output of  $+_n$  can be calculated as

$$net_n = \sum_{j=1}^n \omega_j * h_j \tag{7}$$

where  $h_j (j=1, 2, \dots, n)$  are the inputs to  $+_n$ . Gaussian functions, unipolar sigmoid functions, bipolar sigmoid functions, nonlocal radial coordinates, thin-plate s-spline functions, and general multiquadratics can be adopted as flexible activation functions [2, 14].

For developing the FNT classifier, the following flexible activation function is used:

$$f(a_i, b_i, h) = e^{-\left(\frac{h-a_i}{b_i}\right)^2} \tag{8}$$

where the two adjustable parameters  $a_i$  and  $b_i$  are randomly created as flexible activation function parameters.  $h_j (j=1, 2, 3)$  is the input to  $+_n$ . The output of  $+_n$  is then calculated by

$$out_n = f(a_n, b_n, net_n) = e^{-\left(\frac{net_n-a_n}{b_n}\right)^2} \tag{9}$$

Probabilistic incremental program evolution (PIPE) is selected as a tree-structure-based encoding method with specific instruction sets for fine tuning the parameters encoded in the structure. Starting with the initial set structures and corresponding parameters, PIPE first attempts to improve the structure; then, as soon as an improved structure is found, the parameters of the structure are fine-tuned. Next, PIPE returns to improve the structure again and finds a better structure. The rules' parameters are fine-tuned again [12].

A fitness function arranges the FNT into scalar and real-valued fitness values that reflect the FNT performance according to a given task. In the experiments, the fitness function used for PIPE is given by the RMSE as

$$Fit(i) = \frac{\sum_{j=1}^P (y_1^j - y_2^j)^2}{P} \tag{10}$$

When a satisfactory solution is found or a time limit is reached, the loop ends [14]. The evolved neural tree model is obtained at iteration 28 with the function instruction sets  $I = \{+_2, +_3, +_4, +_5, h_1, h_2, h_3\}$ .

### 2.1.5 Classification with the feature database

The recognition process consists of OPSR preprocessing, OPSR, feature extraction, and classification with the feature database. The first three parts are similar to the training process. However, the last part, “classification with the feature database,” is different. The feature database trained by use of a FNT is adopted for classification purposes. Families of musical instruments and individual musical instruments are used in the classification experiments. The RMSE and recognition rate (Accuracy) are used to verify the classification performance.

## 2.2 Assessment method

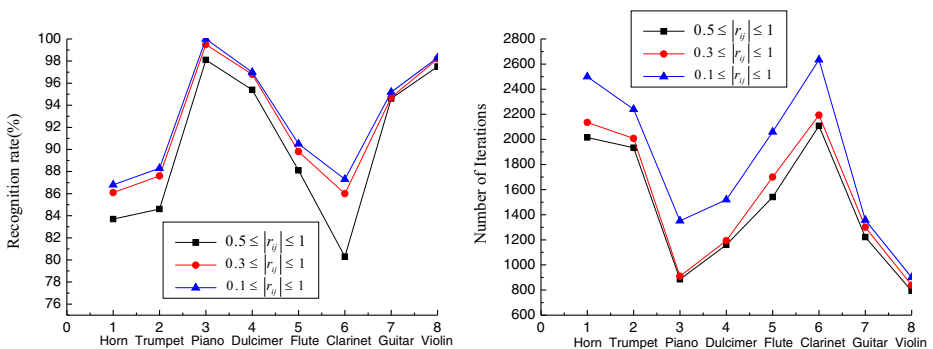
If the choice of the correlation coefficient criterion is not adapted, some error will occur in the experiment. In order to ensure computational efficiency and precision and to minimize the influence of the correlation coefficient, we choose  $0.1 \leq |r_{ij}| \leq 1$ ,  $0.3 \leq |r_{ij}| \leq 1$ , and  $0.5 \leq |r_{ij}| \leq 1$  as the correlation coefficient criteria to test.

As shown in Fig. 4, a larger floor value of the cross correlation criterion results in a higher recognition rate (Accuracy) fewer iterations for the various musical instruments. The dulcimer, piano, violin, and guitar always have relatively high recognition accuracy and a low number of iterations. The recognition accuracy of different instruments are very close when  $0.1 \leq |r_{ij}| \leq 1$  or  $0.3 \leq |r_{ij}| \leq 1$  is selected as the criterion, whereas the number of iterations for  $0.1 \leq |r_{ij}| \leq 1$  is much larger than those for  $0.3 \leq |r_{ij}| \leq 1$  and  $0.5 \leq |r_{ij}| \leq 1$ . Considering the above, we choose  $0.3 \leq |r_{ij}| \leq 1$  as the cross correlation criterion for the second dimension reduction.

## 3 Experiments

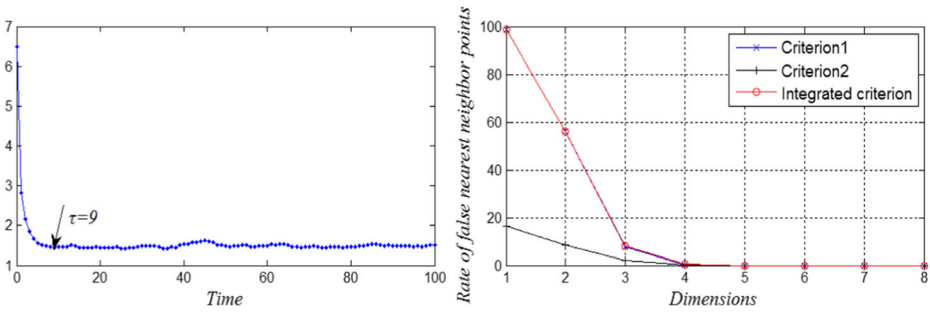
Experiments were conducted to evaluate the performance of the OPSR–PDF–FNT algorithm in musical-instrument classification. We compared its performance with various state-of-the-art methods that were available in the jMIR and Weka open-source software packages, such as the HMM, SVM, bagging, and random forest methods. The RMSE and recognition rates (Accuracy) were used as evaluation metrics to verify the classification performance.

The signals used in the experiment were downloaded in the WAV format from the music database of the University of Iowa. This database contains 761 single samples of 20 types of musical instruments, and all samples are recorded under the same conditions. The duration of



**Fig. 4** Recognition rates (Accuracy) and the number of iterations of various musical instruments for different correlation coefficient criteria





**Fig. 5** Delay time and embedding dimension of a horn sample signal

the samples is more than 4 s [24]. The frequency of the musical-instrument signals is 44,100 Hz. There are 12 instruments belonging to the three main families of musical instruments, and 10 groups of samples per instrument were used in the experiments. After sampling, each group sample takes 5000 samples. A leave-one-out cross validation was adopted to train and test each instrument.

### 3.1 Classification for families of musical instruments

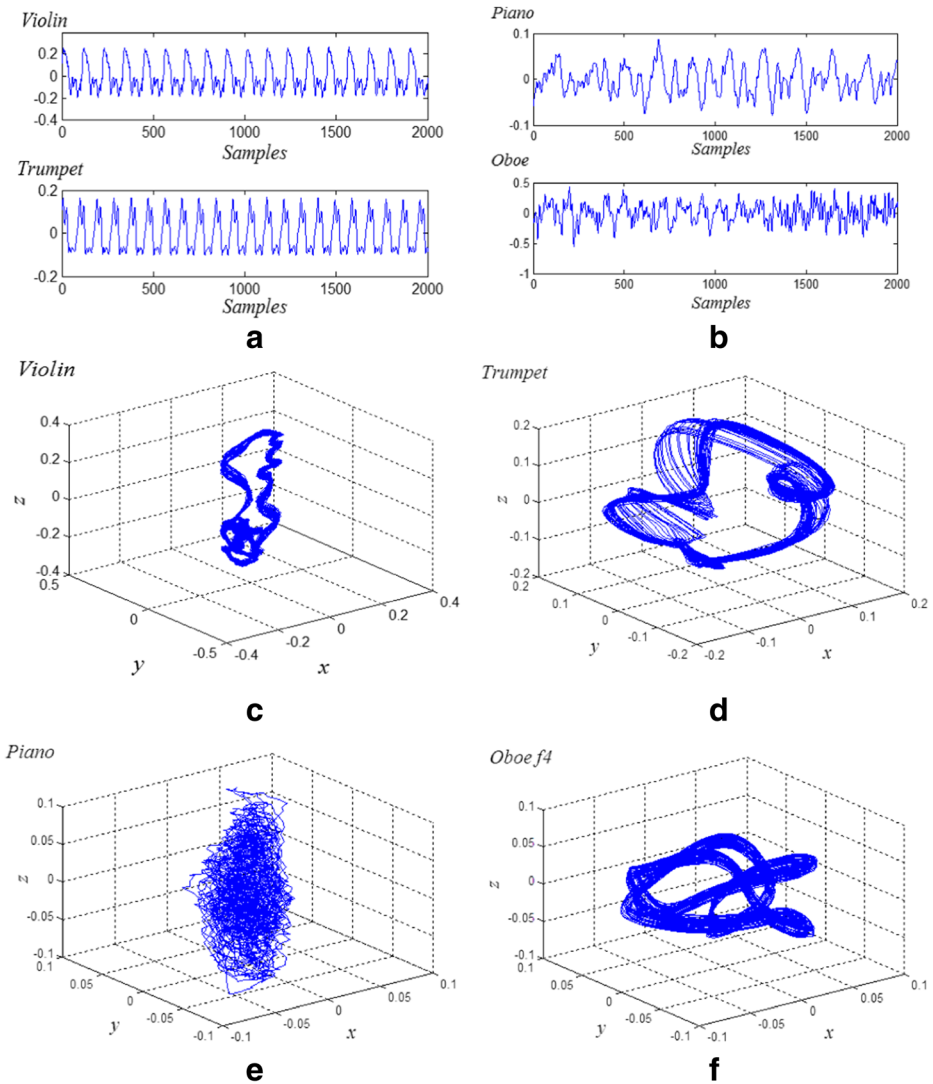
For reconstructing the samples of musical-instrument signals in the phase space, we consider a sample of an instrument with  $\tau$ . This sample is mapped to an  $m$ -dimensional embedded space. Taking a horn sample signal as an example,  $\tau$  and  $m$  can be obtained using the average mutual information method and the FNN method, respectively.

In Fig. 5,  $\tau$  is the corresponding ordinate value 9 of the first zero cross point, and  $m$  is the corresponding abscissa value 5 of the first zero cross point. When we determine the values of  $\tau$  and  $m$ , the musical-instrument sample signals can be reconstructed in a higher phase space. A PCA is used to map the high-dimension reconstructed phase-space vector to a lower dimension space. To improve the performance of the dimension reduction and to shorten the running time of the program, after the first dimension reduction by the PCA, the correlation coefficients between the principal components and the original data of the original musical-instrument signals are computed (see Table 1). In Table 1, taking the horn sample as an example, we find that the third correlation coefficient between the principal components and the original data is 0.1358 and is smaller than 0.3. According to Fig. 4, the third line principal components can be removed. Thus, we realize a dimension reduction by adopting the correlation coefficient criterion.

In order to visualize the dynamic structure of the system, the dimension reduction vector is projected into a three-dimensional space. Figure 6a and b show the sample signals and three-dimensional structure of various families of musical instruments. Figure 6c, d, e, and f show the three-dimensional structure of various families of musical instruments.

**Table 1** Correlation coefficients between the principal components and the original data

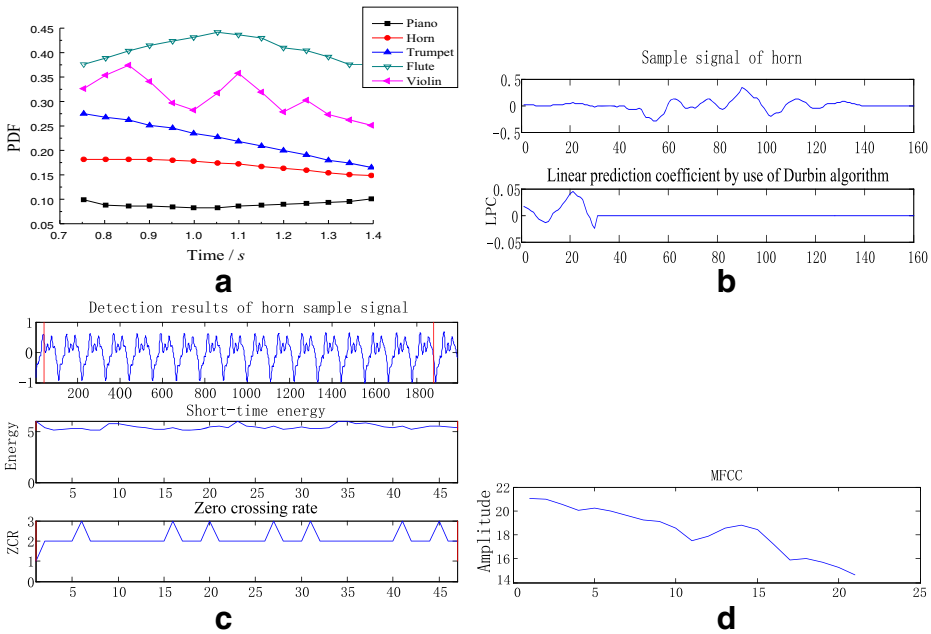
Instruments	Correlation coefficient between principal components and original data					
Horn	0.7661	0.6419	0.1358			
Piano	0.8839	0.4439	0.2293	0.0245		
Trumpet	0.6777	0.7199	0.3124	0.1242	0.0142	
Flute	0.1197	0.7745	0.6210	0.0364	0.0159	0.0114



**Fig. 6** Sample signals and three-dimensional structures of various families of musical instruments

In the time domain, the sample signals are all periodic or periodic-like, and they are not easy to classify. However, various musical instruments have different topological structures in the three-dimensional space. The topological structures of the violin and trumpet show circular orbits, and the track distribution of the trumpet is relatively loose. Though the phase-space tracks of the piano appear messy, the distribution is focused in a relatively stable range. The phase-space tracks of the oboe are similar to a dance ribbon that is twisted and concentrated; however, they present a certain regularity. Furthermore, the phase-space tracks of musical instruments differ depending on the different tones and pitches.

To describe the differences between the various families of musical instruments, the PDF and other traditional features, including the MFCC, short-time energy, zero crossing rate, and linear prediction coefficient are extracted as tone features [3, 9, 17, 19, 22]. Figure 7a depicts



**Fig. 7** Some features of the horn sample signal: the (a) PDF, (b) linear prediction coefficient, (c) short-time energy and zero crossing rate, and (d) MFCC

the PDF values of the sample signals of the piano, horn, trumpet, flute, and violin. We can clearly observe that the flute sample signal has the highest PDF, and the PDF of the piano sample signal is the lowest. The PDF of the horn and trumpet sample signals partly overlap because the horn and trumpet both belong to the brass family. The PDF of the violin is more volatile than the other four instruments. Figure 7b, c, and d show the other tone features of the horn sample signals. All of these features can be combined into a new multidimensional vector. From these multidimensional feature vectors, the instruments in the three families of musical instruments can be distinguished by adopting the PDF and other features.

We use a FNT as a classifier to recognize families of musical instruments. As shown in Fig. 3, the instruction set is  $I = \{+2, +3, +4, +5, h_1, h_2, h_3\}$ . The FNT selects the proper input variables or time lags automatically. In addition, the parameters used for the experiment are listed in Table 2.

**Table 2** Parameters used in the PIPE algorithm for the architecture optimization of the FNT

Population size $PS$	100
Elitist learning probability $P_{el}$	0.01
Learning rate $I_r$	0.01
Fitness constant	0.000001
Overall mutation probability $P_M$	0.4
Mutation rate $m_r$	0.4
Prune threshold $T_p$	0.999
Maximum random search steps	2000
Initial connection weights	rand[-1, 1]
Initial parameters $a_p$ and $b_p$	rand[0,1]

**Table 3** Average recognition rates (Accuracy) of three families of musical instruments (%)

Classifier	FNT	SVM	HMM	Bagging	Random Forest
Feature method					
Tone features	96.2	93.8	94.3	95.4	96.8
PDF	84.6	80.7	82.1		
Comprehensive method	98.3	97.5	97.6		

In comparison to the other models for nonlinear function approximation, the OPSR–PDF–FNT algorithm has better performance with respect to the average recognition rate (Accuracy). In Table 3, the comprehensive method implies the method combining the tone features and PDF.

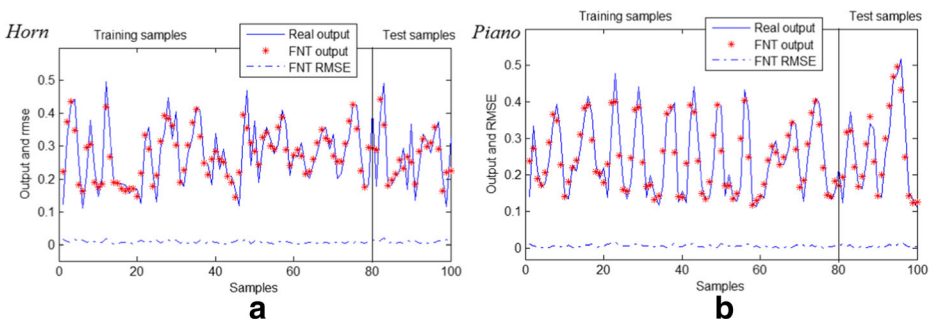
### 3.2 Classification of individual musical instruments

In the experiments, the sample signals of 12 individual musical instruments were reconstructed in a high-dimensional phase space, and the dimension was reduced using OPSR. The features were extracted from the dimension reduction vector by the comprehensive method. The FNT is used to classify 12 individual musical instruments.

Figure 8 shows that the model outputs are close to the real outputs, indicating that the FNT is better for classifying musical instruments. The mean value of the RMSE is 0.000451 for the training data and 0.000573 for the test data.

In Table 4, we adopt the confusion matrix proposed by Stehman in 1997 [28] to show the classification results for 12 musical instruments. Each column of the confusion matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The results enclosed in parentheses are the recognition rates (Accuracy) achieved using the OPSR–tone-feature–FNT method, whereas those outside the brackets are the means of the recognition rates (Accuracy) achieved using the OPSR–PDF–FNT method. The diagonal data express the recognition accuracy for each musical instrument.

From Table 4, we find that the recognition rates (Accuracy) of the OPSR–PDF–FNT method are superior to those of the OPSR–tone-feature–FNT method. In addition, the experiment results for keyboard instruments and string instruments are obviously better than those for wind instruments. Taking the violin as an example, the correct recognition rate (Accuracy) is 95.8 % when adopting the OPSR–PDF–FNT method. However, the violin is classified as a guitar (3.9 %) and as a flute (0.3 %). The mean recognition rates (Accuracy) for individual

**Fig. 8** FNT model outputs and RMSE

**Table 4** Recognition rates (Accuracy) for 12 musical instruments (%)

	Horn	Trumpet	Piano	Accordion	Organ	Flute	Clarinet	Oboe	Sax	Guitar	Violin	Cello
Horn	86.9 (85.1)	5.3 (6.5)	0	0	0	1.9 (2.2)	4.8 (4.8)	1.1 (1.4)	0	0	0	0
Trumpet	2.4 (3.3)	93.4 (91.3)	0	0	0	0	2.3 (2.8)	0	1.9 (2.6)	0	0	0
Piano	0	0	100 (96.5)	0	0	0	0	0	0	0	0	0
Accordion	0	0	1.1 (1.7)	98.2 (97.6)	0.3 (0.3)	0	0	0.4 (0.4)	0	0	0	0
Organ	0	0	1.6 (3.5)	1.1 (3.3)	96.3 (91.9)	0	0	1 (1.3)	0	0	0	0
Flute	0.4 (1.1)	0	0	0	0	90.7 (86.5)	4.8 (6.4)	0	3.6 (4.9)	0	0.5 (1.1)	0
Clarinet	0	0	0	0	0	5.5 (7.4)	86.3 (80.3)	0	7.3 (8.1)	0	0	0.9 (4.2)
Oboe	0	0	0	0	0	3.7 (4.5)	4.3 (5.9)	90.5 (85.6)	0.9 (1.5)	0	0	0.6 (1.3)
Sax	2.7 (3.6)	0	0	0	0	3.3 (3.8)	3.9 (4.2)	0.4 (2.1)	89.7 (86.3)	0	0	0
Guitar	0	0	1.1 (3.8)	0	0	0	0	0	0	94.3 (87.9)	3.4 (3.5)	1.2 (1.6)
Violin	0	0	0	0	0	0.3 (2.1)	0	0	0	3.9 (4.6)	95.8 (92.5)	0
Cello	0	0	0	0	0	0.1 (3.7)	0	0	0	0.6 (0.8)	1.7 (1.7)	97.6 (93.8)

musical instruments increase to 93.3 % by using the proposed method, and they are higher than 89.6 % using the OPSR–tone-feature–FNT method.

## 4 Conclusion

In this paper, a musical-instrument classification method by combining OPSR with the PDF and FNT was proposed to classify individual musical instruments and families of musical instruments. It was shown that the OPSR–PDF–FNT method has several desirable properties for musical-instrument signal classification:

First, as per nonlinear dynamic theory, PSR maps a one-dimensional time series of a musical instrument to an  $m$ -dimensional embedded space. The multidimensional PSR results maintain the topological characteristics of the original time series and contain more dynamic information.

Second, a PCA and correlation coefficient are used to optimize the PSR method. The multidimensional PSR results of different musical-instrument signals are extracted as the main components, and the dimensionality is reduced by OPSR.

Third, in order to describe the differences between various musical instruments, we use the PDF and other traditional features to construct a comprehensive method to extract the features of musical instruments.

Fourth, a FNT is good at evolving an approximating model of the static nonlinear system. With this advantage, the FNT is adopted as a classifier to tackle the variability in musical-instrument signals and improve the adaptive ability of various target classification problems.

The proposed algorithm was compared with the OPSR–PDF–SVM, OPSR–PDF–HMM, and OPSR–tone-feature–FNT algorithms. The simulated results showed that proposed algorithm outperforms the other three algorithms with higher recognition rates (Accuracy) and lower RMSEs.

There are two directions to take this work in the future. The first is to build a more concise model of OPSR to better reduce the dimension of the PSR vectors. Second, the addition of a recurrence analysis method to OPSR could aid in extracting more characteristics of the musical-instrument signals.

**Acknowledgments** Funding for this work was supported by the National Natural Science Foundation of China (NO. 61301250, NO. 61401298), Program for the Outstanding Innovative Teams of Higher Learning Institutions of Shanxi (NO. [2015]3), Project of Shanxi Scholarship Council of China (NO. 2014-060), Doctoral Program of Taiyuan University of Science and Technology (NO. 20152003), and Project for “131” Talented Person Project of Higher Learning Institutions of Shanxi (NO. [2016]).

## References

1. Agostini G (2003) Musical instrument timbres classification with spectral features. *EURASIP J Appl Signal Process* 1:5–14
2. Bao W, Chen Y, Wang D (2014) Prediction of protein structure classes with flexible neural tree. *Bio-Med Mater Eng* 24(6):3797–3806

3. Barbedo JGA, Tzanetakis G (2011) Musical instrument classification using individual partials. *IEEE Trans Audio Speech Lang Process* 19(1):111–122
4. Benetos E, Kotti M, Kotropoulos C (2007) Large scale musical instrument identification. Paper presented at the 4th Sound and Music Computing Conference 11: 13
5. Bhalke DG, Rao CBR, Bormane DS (2014) Musical instrument classification using higher order spectra. *Signal Processing and Integrated Networks (SPIN)*, 2014 International Conference on. IEEE 40–45
6. Bouaziz S, Dhahri H, Alimi AM et al (2013) A hybrid learning algorithm for evolving flexible beta basis function neural tree model. *Neurocomputing* 117:107–117
7. Chen Y, Abraham A (2009) *Tree-structure based hybrid computational intelligence: Theoretical foundations and applications*. Springer Science & Business Media 2:39–96
8. Chen Y, Abraham A, Yang B (2006) Feature selection and classification using flexible neural tree. *Neurocomputing* 70(1):305–313
9. Deng JD, Simmermacher C, Cranefield S (2008) A study on feature analysis for musical instrument classification. *IEEE Trans Syst Man Cybern B Cybern* 38(2):429–438
10. Eronen A (2001) Comparison of features for musical instrument recognition. *Applications of Signal Processing to Audio and Acoustics, 2001 I.E. Workshop on the. IEEE* 19–22
11. Guo Y, Huang S, Li Y (2012) Single-mixture source separation using dimensionality reduction of ensemble empirical mode decomposition and independent component analysis. *Circuits, Systems, and Signal Processing* 31(6):2047–2060
12. Guo Y, Wang Q, Huang S, Abraham A (2012) Flexible neural trees for online hand gesture recognition using surface electromyography. *J Comput* 7(5):1099–1103
13. Guo Y, Huang S, Li Y, Ganesh RN (2013) Edge effect elimination in single-mixture blind source separation. *Circuits, Systems, and Signal Processing* 32(5):2317–2334
14. Guo Y, Wang Q, Huang S, Abraham A (2014) Hand gesture recognition system using single-mixture source separation and flexible neural trees. *J Vib Control* 20(9):1333–1342
15. Hess S, Kitaura FS (2016) Cosmic flows and the expansion of the local universe from non-linear phase-space reconstructions. *Mon Not R Astron Soc* 456(4):4247–4255
16. Hong M, Wang D, Wang Y et al (2016) Mid- and long-term runoff predictions by an improved phase-space reconstruction model. *Environ Res* 148:560–573
17. Joder C, Essid S, Richard G (2009) Temporal integration for audio classification with application to musical instrument classification. *IEEE Trans Audio Speech Lang Process* 17(1):174–186
18. Jolliffe I (2002) *Principal component analysis*. John Wiley & Sons, Ltd
19. Kazi FI, Bhalke DG (2014) Musical instrument classification using higher order spectra and hierarchical taxonomies. *Int J Comput Appl* 107(17):17–22
20. Koulaouzidis G, Das S, Cappiello G et al (2015) Prompt and accurate diagnosis of ventricular arrhythmias with a novel index based on phase space reconstruction of ECG. *Int J Cardiol* 182:38–43
21. McKay C (2010) *Automatic music classification with jMIR*. Doctoral dissertation, McGill University
22. Mison R, Rosli N, Manaf NA, Halim HA (2014) Music emotion classification (mec): Exploiting vocal and instrumental sound features. *Recent Advances on Soft Computing and Data Mining*. Springer International Publishing, pp 539–549
23. Patil K, Elhilali M (2015) Biomimetic spectro-temporal features for music instrument recognition in isolated notes and solo phrases. *EURASIP JASMP* 1:1–13
24. Rui R, Bao C (2012) The musical instrument classification algorithm based on nonlinear dynamics. *Acta Electron Sin* 7:032
25. Stehman SV (1997) Selecting and interpreting measures of thematic classification accuracy. *Remote Sens Environ* 62(1):77–89
26. Takens F (1981) Detecting strange attractors in turbulence. *Dynamical systems and turbulence*, Warwick 1980. Springer, Berlin Heidelberg, pp 366–381
27. Wang Y, Wang J, Wei X (2015) A hybrid wind speed forecasting model based on phase space reconstruction theory and Markov model: a case study of wind farms in northwest China. *Energy* 91:556–572
28. Xu T, Wang Y, Zhang Z (2013) Pixel-wise skin colour detection based on flexible neural tree. *IET Image Process* 7(8):751–761
29. Xu B, Jacquir S, Laurent G et al (2014) Analysis of an experimental model of in vitro cardiac tissue using phase space reconstruction. *Biomed Signal Process Control* 13:313–326
30. Yang B, Chen Y (2016) Somatic mutation detection using ensemble of flexible neural tree model. *Neurocomputing* 179:161–168



**Yina Guo** received the Ph.D. degree from Taiyuan University of Science and Technology, Taiyuan, Shanxi, P.R.China. She joined the Department of Electronic and Communication, Taiyuan University of Science and Technology as an associate professor from 2012. Her research interests focus on single channel blind source separation, musical instrument recognition and hand gesture recognition. She has taken charge of a General Program of National Natural Science Foundation of China and several scientific projects of Shanxi province. She has obtained 6 patents and 4 software copyrights of China. In addition, she published more than 30 academic papers and 2 books.



**Qijia Liu** received the B. Sc. degree from North University of China, Taiyuan, Shanxi, P.R.China. His research interests focus on single channel blind source separation and hand gesture recognition.





**Anhong Wang** received the Ph.D. degree from Beijing Jiaotong University, Beijing, P.R.China. She joined the Department of Electronic and Communication, Taiyuan University of Science and Technology as a professor from 2009. Her research interests focus on Video/image coding and transmission. She has taken charge of several scientific projects of National Fund and Shanxi province. She has published more than 100 academic papers and obtained 10 patents of China.



**Chaoli Sun** received the Ph.D. degree from Taiyuan University of Science and Technology, Taiyuan, Shanxi, P.R.China. She joined the Department of Computer Science and Technology, Taiyuan University of Science and Technology as an associate professor from 2011. Her research interests focus on swarm intelligent optimization algorithm. She has taken charge of several scientific projects of National Fund and Shanxi province. She has published more than 30 academic papers and 1 book.



**Wenyan Tian** received the Ph.D. degree from Sichuan University, Chengdu, P.R.China. She joined the Department of Electronic and Communication, Taiyuan University of Science and Technology as a professor from 2009. Her research interests focus on microwave nonthermal effects, a General Program of National Natural Science Foundation of China and several scientific projects of Shanxi province. She has published more than 20 academic papers and obtained 2 patents and 3 software copyrights of China.



**Ganesh R. Naik** received the PhD degree in the area of Electronics Engineering, specialised in Biomedical Engineering and Signal processing from RMIT University, Melbourne, Australia, in 2009. He is currently Chancellor's Post Doctoral Research Fellow at Faculty of Engineering and Information Technology (FEIT), UTS. As an early career researcher, he has edited 10 books, authored more than 80 papers in peer reviewed journals, conferences, and book chapters over the last 7 years. His research interests include EMG signal processing, Pattern recognition, Blind Source Separation (BSS) techniques, Biomedical signal processing, Human Computer Interface (HCI) and Audio signal processing.



**Dr. Abraham's** research and development experience includes more than 23 years in the industry and academia. He received the Ph.D. degree in Computer Science from Monash University, Melbourne, Australia. He works in a multi-disciplinary environment involving machine (network) intelligence, cyber security, sensor networks, Web intelligence, scheduling, data mining and applied to various real world problems. He has given more than 60 conference plenary lectures/tutorials and invited seminars/lectures in over 75 Universities around the globe. He is an author/co-author of 900+ publications and some of the works have also won best paper awards at International conferences.

## Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

[onlineservice@springernature.com](mailto:onlineservice@springernature.com)