Audio Classification Algorithm Based on Nonlinear Characteristics Analysis

Liyan Zhang, Changchun Bao, Xin Liu, Xingtao Zhang, Feng Bao and Bing Bu
Speech and Audio Signal Processing Laboratory, School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing
E-mail: zhangliyan1999@emails.bjut.edu.cn, baochch@bjut.edu.cn, liuxin0930@emails.bjut.edu.cn, taozi@emails.bjut.edu.cn baofeng0620@gmail.com, nono_bubing@emails.bjut.edu.cn

***Abstract*—In this paper, an audio classification algorithm based on nonlinear characteristics analysis is proposed. According to the nonlinearity of audio signals, the phase space reconstruction technology of nonlinear dynamics, the recurrence plot and the recurrence quantification analysis are introduced to audio characteristics analysis, and the audio signals are classified into four types: noise-like, transient, harmonic-like and mixed signal. Test results indicate that the proposed algorithm has good classification accuracy.**

I. INTRODUCTION

As a focused research on audio signal processing and analysis nowadays, audio classification has been widely used in the field of audio coding, audio retrieval and audio enhancement. Due to the inherent characteristics of audio information, the audio classification techniques play an important role for the structured audio processing.

Conventional audio classification algorithms are commonly based on content and classify audio into silence, clean speech, music and noisy speech [1]. Based on these classification algorithms, each type of audio is processed by different methods. However, audio signal is still processed by a unified model for speech and music signal, which is ineffective for harmonic and transient information [2][3].

In this paper, an audio classification algorithm based on nonlinear characteristics analysis is proposed. According to the nonlinearity of audio signals, the phase space reconstruction technology of nonlinear dynamics is introduced, the recurrence plot and the recurrence quantification analysis are used to analyze audio characteristics, and audio signals are classified into noise-like, transient, harmonic-like and mixed signal. This lays the foundation for audio signal processing.

This paper is organized as follows. After validating the nonlinearity of audio time series, the recurrence plot and the recurrence quantification analysis of nonlinear dynamics are given in Section Ⅱ, whereas the main block of the proposed classification algorithm is described in Section Ⅲ. The classification accuracy evaluation is presented in Section Ⅳ and the Section Ⅴ will draw the conclusions.

II. RECURRENCE PLOT AND RECURRENCE QUANTIFICATION ANALYSIS

A. Recurrence Plot

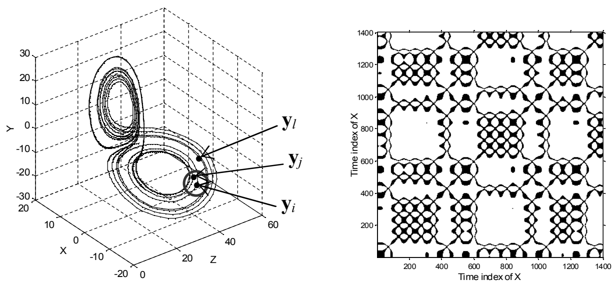
The recurrence of states in the nature has been known for a long time and has also been discussed in early publications by Monk in 1939[4]. Usually, a phase space whose dimension is less than 4 is allowed to be pictured. The phase space of a higher dimension can only be visualized by projection into the two or three dimensional sub-spaces. Eckmann [5] et al. have introduced a tool which enables us to investigate the *m*-dimension phase space trajectory through a two-dimension representation of its recurrences. This representation is called recurrence plot (RP). RP can be mathematically expressed as

$$R_{i,j} = \Theta(\varepsilon_i - \left\| \mathbf{y}_i - \mathbf{y}_j \right\|), \mathbf{y}_i \in \Re^m, i, j = 1, 2, \cdots, N \qquad (1)$$

where N is the number of the considered states $\mathbf{y}_i$, ε_i is a predefined threshold, $\|\bullet\|$ is the norm (e.g. the Euclidean norm) and $\Theta(\bullet)$ is the Heaviside function, which is defined as

$$\Theta(z) = \begin{cases} 0 & z < 0 \\ 1 & z \geq 0 \end{cases} \qquad (2)$$

The recurrence of a state at time i with regard to time j is represented as a two-dimension square matrix with black and white points, where the horizontal and vertical axes on recurrence plot are all the time axes.



(a) A phase space trajectory (b) Recurrence plot

Fig.1 A phase space trajectory and its recurrence plot of Lorenz system

Fig.1 (a) is an example of the phase space trajectory of the three-dimension Lorenz system and Fig.1 (b) is the corresponding RP. The gray circle in Fig.1 (a) is a region whose centre is $\mathbf{y}_i$ and radius is ε. The vectors $\mathbf{y}_i$ and $\mathbf{y}_j$ of the trajectory falling into this region are represented as a black point in terms of coordinate (i, j) in the RP. On the contrary, the vectors $\mathbf{y}_i$ and $\mathbf{y}_l$ not falling in this region are represented as a white point in terms of coordinate (i, l) in the RP.

B. Recurrence Quantification Analysis

After intuitively describing the recurrence characteristics by RP, it is necessary to analyze it quantificationally. Zbilut and Webber have developed a tool named as recurrence quantification analysis (RQA) [6], which quantifies the mentioned structures in the RP. It has been extended with some new measures of the complexity by Marwan [7]. Next, we will introduce these RQA measures used in this paper.

Recurrence rate R_R: the percentage of recurrence points (black points) in the RP and it is defined by

$$R_R = \frac{1}{N^2}\sum_{i,j=1}^{N} R_{i,j} \tag{3}$$

It reveals the occurrence probability of the similar states in the system. The larger the R_R is, the higher the probability that system trajectories fall into the same phase space regions is.

Determinism R_D: the percentage of recurrence points forming the diagonal structures with a 45° angle in all recurrence points and it is defined by

$$R_D = \sum_{l=l_{\min}}^{N} lP(l) \Big/ \sum_{i,j=1}^{N} R_{i,j} \tag{4}$$

where $P(l)$ is the probability of the diagonal lines with the length of l and $l_{\min}$ is the minimal length of the diagonal lines. The fact is that the random system could hardly produce the diagonal structure, whereas it is easier to produce a diagonal structure for the deterministic system.

The mean of diagonal line length L_{mean} is given by

$$L_{mean} = \sum_{l=l_{\min}}^{N} lP(l) \Big/ \sum_{l=l_{\min}}^{N} P(l) \tag{5}$$

The length of the longest diagonal line $L_{\max}$ is given by

$$L_{\max} = \max(\{l_i; i = 1, 2, \cdots, N_l\}) \tag{6}$$

where N_l is the number of diagonal lines in the RP.

Laminarity R_L: the percentage of recurrence points forming the vertical structures in all recurrence points and it is defined by

$$R_L = \sum_{v=v_{\min}}^{N} vP(v) \Big/ \sum_{i,j=1}^{N} R_{i,j} \tag{7}$$

where $P(v)$ is the probability of the vertical lines with the length of v and $v_{\min}$ is the minimal length of the vertical lines. Here, the R_L is a measure of the amount of vertical structures in the whole RP and represents the occurrence of laminar states in the system.

The mean V_{mean} of the vertical lines length is defined by

$$V_{mean} = \sum_{v=v_{\min}}^{N} vP(v) \Big/ \sum_{v=v_{\min}}^{N} P(v) \tag{8}$$

The length of the longest vertical line: $V_{\max}$ is given by

$$V_{\max} = \max(\{v_i; i = 1, 2, \cdots, N_v\}) \tag{9}$$

where N_v is the number of vertical lines in the RP.

III. AUDIO CLASSIFICATION

Generally, the time-frequency characteristics of audio signals are commonly analyzed in the view of auditory perception in advance. Some related researches find that the nonlinearity of audio signals is typical in time domain, since it is affected by the different states of sound source and the surroundings of sound wave transmission [8]. Now, the dynamical invariant, maximum Lyapunov exponent has been used to determine the nonlinearity of audio signals [9]. In this case, we can analyze audio signals on the basis of the nonlinear characteristics analysis of nonlinear dynamics.

A. Phase Space Reconstruction of Audio Signals

According to the phase space reconstruction method proposed by Takens [10], the one-dimension nonlinear time series of audio signal $\mathbf{x}=(x_1,x_2,x_3,\ldots,x_M)^{\mathrm{T}}$ can be reconstructed to a state matrix $\mathbf{Y}$ which is given by

$$\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \cdots \mathbf{y}_N) = \begin{pmatrix} x_1 & x_2 & \cdots & x_{M-(m-1)\tau} \\ x_{1+\tau} & x_{2+\tau} & \cdots & x_{M-(m-2)\tau} \\ x_{1+2\tau} & x_{2+2\tau} & \cdots & x_{M-(m-3)\tau} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1+(m-1)\tau} & x_{2+(m-1)\tau} & \cdots & x_M \end{pmatrix} \tag{10}$$

where M is the number of points of one-dimension nonlinear time series $\mathbf{x}$, τ is the embedding time delay, m is the embedding dimension, $N=M-(m-1)\tau$ is the number of points for the phase space trajectory and $\mathbf{y}_i$ is one of points in m-dimension phase space which represents the i^{th} state of the system. And the m-dimension phase space trajectory is composed of the states $\{\mathbf{y}_i, i = 1, 2, 3, \ldots, N\}$.

Before plotting the RP of time series of audio signals, there are three parameters to be predefined: the embedding time delay τ, the embedding dimension m and the threshold distance ε. In this paper, the embedding time delay τ and the embedding dimension m are represented by autocorrelation method and false nearest neighbor method [2], respectively.

It is important to determine threshold distance ε_i. There is not a unified method to calculate it for the moment. In this paper, we will predefine ε_i by the recurrence rate. It is known that the recurrence rate reveals the density of recurrence points and the convergence of trajectories in the phase space. Using (1) and (3), the recurrence rate is given by

$$R_R = \frac{1}{N^2}\sum_{i,j=1}^{N} \Theta(\varepsilon_i - \|\mathbf{y}_i - \mathbf{y}_j\|) \tag{11}$$

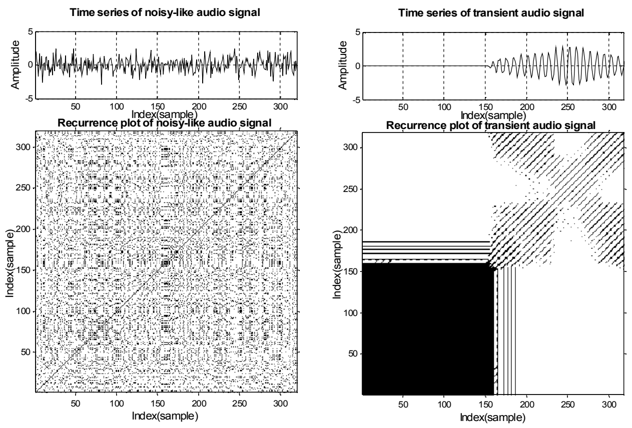
For a frame of audio signal, after reconstructing the phase space, the relationship between threshold distance ε and recurrence rate R_R is given by Table I. And σ is used to represent standard error of the audio signal in time domain.

TABLE I
RELATIONSHIP BETWEEN THRESHOLD DISTANCE AND RECURRENCE RATE

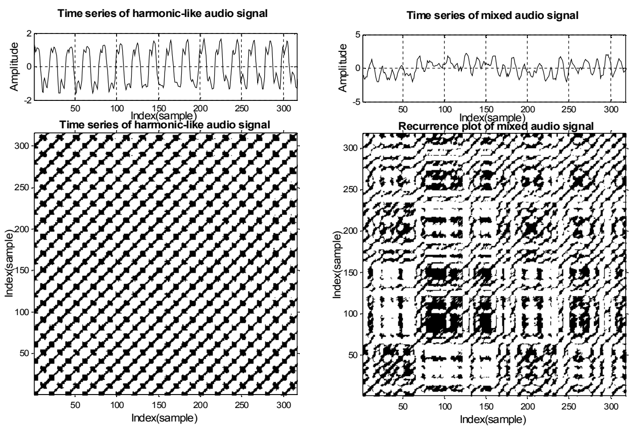
ε	0.3σ	0.5σ	1.0σ	1.3σ	1.5σ
R_R	1.7%	5.7%	20.7%	30.0%	37.5%
ε	2.0σ	2.5σ	3.0σ	3.5σ	3.8σ
R_R	59.0%	82.3%	95.7%	99.5%	99.9%

As depicted in Table I, threshold distance and recurrence rate are one-to-one relationship when the audio signal is known. Recurrence rate is increasing along with the increase of threshold distance. Therefore, the threshold distance can be adjusted according to recurrence rate to exactly describe the characteristics of recurrence points in the RP. In this case, the recurrence rate is further used to determine recurrence characteristic of audio signals, combining with the embedding delay time and the embedding dimension.

B. Nonlinear Characteristics Analysis of Audio Signals



(a) RP of noisy-like audio signal (b) RP of transient audio signal



(c) RP of harmonic-like audio signal (d) RP of mixed audio signal

Fig. 2 Recurrence plots of audio signals

According to the global topological structure of the RP, the audio signals are classified into noise-like, transient, harmonic-like and mixed signal. The RP of noise-like audio signal is shown in Fig. 2(a). There are numerous isolated recurrence points uniformly distributed and very few diagonal lines or vertical lines in Fig. 2(a). This reflects the noisy characteristic of this frame. The RP of transient audio signal is depicted in Fig. 2(b), the abrupt changes in the dynamics as well as extreme events cause white areas or bands in the RP which is easy to find and distinguish. The diagonal lines, the checkerboard structures and the periodic recurrent structures exist in the RP of harmonic-like audio which is shown in Fig. 2(c). For this quasi-periodic system, the distances between the diagonal lines are almost equal. It is easily to find its periodicity. The RP of mixed audio signal is depicted in Fig. 2(d), there are less isolated recurrence points comparing with noise-like signal, less vertical lines comparing with transient signal, and less diagonal lines comparing with harmonic-like signal. It is mixed with noise-like, transient and harmonic-like characteristics.

The RP intuitively represents the *m*-dimension phase space trajectory of the dynamical system through a two-dimension figure, which contains the evolving trend of phase space trajectory and reveals the variation regularity of the internal structure for the system. In order to describe the regularity quantitatively, it is necessary to introduce recurrence quantification analysis.

In the RQA, R_R reveals the density of the recurrence points and the convergence of trajectories in the phase space. The higher value of R_R implies a higher probability of the occurrence for the same state in the system. R_D, L_{mean} and L_{max} reveal the periodicity of trajectory, and the larger values imply a longer time span of the occurrence for a similar dynamics in the system. R_L, V_{mean} and V_{max} reveal the change speed of the states and reflect the stability of the system.

The audio signals used for analysis in this paper are from MPEG. The audio signals of 200s are sampled at 16 kHz. The length of analysis frame is 20ms. Then, the recurrence characteristics of four types are analyzed. L_{mean} and V_{mean} of noise-like signal is less than 7, and L_{max} and V_{max} is less than 60, which reflect that most of recurrence points of noise-like signal are isolated and the structures of diagonal lines and the vertical lines are few. The R_D and R_L of transient signal are larger than 80%, the V_{mean} is larger than 7 and the V_{max} is larger than 60, which demonstrate that the abrupt changes exist in a certain period of time. The R_D and R_L of harmonic-like signals are larger than 85%, which is the representation of certainty, in the meantime, L_{max} is larger than 160, which is the representation of harmonic.

C. Classification Algorithm of Audio Signals

The thresholds of four types of audio signals are shown in Table Ⅱ according to the RQA. R_R is defined as 30%, which means the percentage of recurrence points in the RP of audio signals is 30%. First, we will analyze whether a signal belongs to the noise-like type. If not, we continue to analyze whether the transient type will occur, and next the harmonic-like type is analyzed. Finally, the signal which does not belong to the noise-like, transient or harmonic-like type is considered as a mixed type audio signal.

TABLE Ⅱ
THRESHOLDS OF AUDIO CLASSIFICATION ALGORITHM

	R_R	R_D	L_{mean}	L_{max}	R_L	V_{mean}	V_{max}
Noise-like	30%		<7	<60		<7	<60
Transient	30%	>80%			>80%	>7	>60
Harmonic-like	30%	>85%		>160	>85%		
Mixed	30%						

IV. EVALUATION AND TEST RESULTS

In order to evaluate the performance of the proposed algorithm, the data used for test are selected from MPEG (containing drum, violin, harmonica, sinfonia, popular song and some percussion instruments) and SQAM (containing claves, side drum, triangle and gong), respectively.

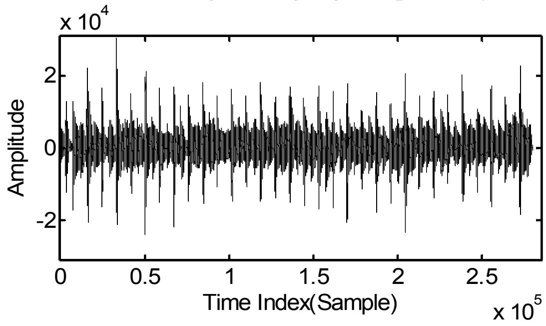


Fig. 3 Waveform of drum signal

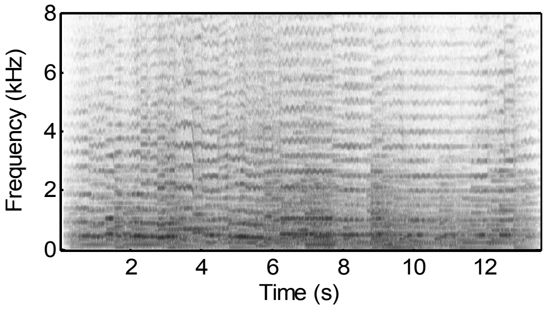


Fig. 4 Spectrogram of violin signal

Fig. 3 shows the waveform of drum signal. Its abrupt changes of amplitude obviously reflect the transient characteristics. Fig. 4 shows the spectrogram of violin signal which illuminates its harmonic characteristics intuitively. The results of classification based the proposed algorithm is shown in Table Ⅲ.

TABLE Ⅲ
THE RESULT OF CLASSIFICATION OF DRUM AND VIOLIN

	Noise-like	Transient	Harmonic-like	Mixed
Drum	0.23%	46.06%	52.34%	1.37%
Violin	0.88%	4.12%	89.85%	5.15%

Table Ⅲ shows that, the percentage of transient type for drum is 46.06%, which reflects its abrupt changes of amplitude, and the percentage of harmonic type for violin is 89.85%, which corresponds to the harmonic characteristics of spectrum. The results of classification for drum and violin show that the classification algorithm is effective.

TABLE Ⅳ
RESULT OF CLASSIFICATION OF FOUR TYPES AUDIO SIGNALS

Audio / Type	Noise-like	Transient	Harmonic-like	Mixed
Noise-like	87.5%	0	2%	6%
Transient	2.75%	84%	2%	3.5%
Harmonic-like	2.25%	10%	90.25%	8%
Mixed	7.5%	6%	5.75%	82.5%

In addition, some audio signals of four types with the length of 32s are selected for test. The results are shown in Table Ⅳ, and their classification accuracies are larger than 82%, which also reflects the efficiency of the proposed algorithm.

V. CONCLUSIONS

An audio classification algorithm based on nonlinear characteristics analysis is proposed in this paper. The spectrum of audio is re-described by means of phase space reconstruction of nonlinear dynamics, and the phase space trajectory is investigated through a two-dimension recurrence plot. The variation law of the internal structure for the system in the RP is described by the recurrence quantification analysis. Audio signals are classified into noise-like, transient, harmonic-like and mixed signal. Test results demonstrate that the proposed algorithm has a good performance.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 61072089, Grant No. 60872027), the Beijing Municipal Natural Science Foundation (Grant No. 4082006), the Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality, the Doctoral Research Initiation Foundation Project of Beijing University of Technology and the 9th Postgraduate Science Foundation of Beijing University of Technology (ykj-2011-4910).

REFERENCES

[1] L. Lu, H. J. Zhang and H. Jiang, “Content analysis for audio classification and segmentation,” *IEEE Trans. Speech Audio Process. USA*, vol. 10, pp. 504-516, October 2002.

[2] X. Liu, C. C. Bao, M. S. Jia and Y. T. Sha, “A harmonic bandwidth extension based on Gaussian mixture model,” *Int Conf Signal Process Proc. Beijing*, pp. 474-477, 2010.

[3] Y. T. Sha, C. C. Bao, M. S. Jia and X. Liu, “High frequency reconstruction of audio signal based on chaotic prediction theory,” *ICASSP 2010. Dallas*, pp. 381-384, 2010.

[4] A. T. Monk, A. H. Compton, “Recurrence phenomena in cosmic-ray intensity,” *Rev. Mod. Phys.* vol. 11, pp. 173-179, 1939.

[5] J. P. Eckmann, S. O. Kamphorst and D. Ruelle, “Recurrence plots of dynamical systems,” *Europhys. Lett. Switzerland*, vol. 4, pp. 973-977, November 1987.

[6] J. P. Zbilut and C. L. Webber Jr., “Embeddings and delays as derived from quantification of recurrence plots,” *Phys. Lett. A. Netherlands*, vol. 171, pp. 199-203, December 1992.

[7] N. Marwan, *Encounters with neighbours-current developments of concepts based on recurrence plots and their applications*, University of Potsdam, 2003.

[8] A. Kumar, S. K. Mullick, “Nonlinear dynamical analysis of speech,” *J. Acoust. Soc. Am. USA*, vol. 100, pp. 615-629, 1996.

[9] X. Liu, C. C. Bao, M. S. Jia and Y. T. Sha, “Nonlinear bandwidth extension based on nearest-neighbor matching,” *APSIPA ASC - Asia-Pac. Signal Inf. Process. Assoc. Annu. Summit Conf. Singapore*, pp. 169-172, 2010.

[10] F. Takens, “Detecting strange attractors in turbulence,” *Lecture Notes in Math.*, vol. 898, pp. 366-381, 1981.