

COVER PAGE

Document downloaded by @DAEL

Sat Nov 28 09:28:13 2026
For personal use

When automatic English translation is provided, only the original document is authentic.
The EAA cannot be held responsible of any translation error

Bibliographical reference

Study on the Correlation Between the Temporal Factors and the Soundscape Excerpts Assessment, G. B. Chen and Y. Z. Zhao, Acta Acustica **vol. 102** (Number 4), 2016, pp. 663-674

DOI
https://doi.org/10.3813/AAA.918983

Acta Acustica united with Acustica
Vol. 102 (2016) 663 – 674
DOI 10.3813/AAA.918983

Study on the Correlation Between the Temporal Factors and the Soundscape Excerpts Assessment

G. B. Chen, Y. Z. Zhao
State Key Laboratory of Subtropical Building Science, School of Architecture, South China University of Technology, No. 381, Wushan Road, Tianhe District, Guangzhou, Guangdong Province, China.
brucelee2500@163.com

Summary
Temporal parameters were extracted from the temporal fluctuation spectra and the autocorrelation functions of short-time intensity and pitch sequences of 81 soundscape excerpts, which were then classified into nine clusters. Several representative samples in each cluster were selected and were reproduced for 50 subjects who were instructed to assess a set of attributes of the sounds that they heard. The first four principle components of the data were examined, and their correlations with temporal parameters were then studied. The resulting clustering can correspond to the data of the soundscape assessment to some extent as sample distribution was displayed in principle component plots, and some of the parameters were proven to be significantly correlated with the assessment according to the Pearson's and Spearman's coefficients and a stepwise multiple linear regression analysis. However, the data variance that was explained by the parameters was not large, and it was assumed that the effect of temporal factors on soundscape assessment was complex and cannot be resolved only by linear correlation analysis.

PACS no. 43.50.Rq

1. Introduction

Soundscape research is a relatively new field of science that is based on a listener-centered model, that concentrates on a method of initiatively creating a sonic environment that benefits and has a positive effect on people's lives [1]. However, multiple dimensions are known to be involved in the perception of environmental sound, including physical, psychoacoustic, cognitive and personal factors [2, 3]. The temporal aspect, which is a special way in which listeners perceive, experience and comprehend the sonic environment, has attracted relatively less attention to date.

The earliest publication concerned with tempo in the soundscape is *The Tuning of the World*, which was written by R. M. Schafer, the father of acoustic ecology [4]. As a composer, he treated the world as a macrocosmic musical composition and emphasized the infinite variety of rhythms in the universe. The human body, on the other hand, has several audible metrical divisions such as the heart, breath and foot and even the conservational actions of the nervous system, which must be the guide against which we arrange the rhythms of the environment, including the soundscape.

Received 30 October 2015,
accepted 13 May 2016.

Numerous field surveys and laboratory listening tests on environmental sounds were conducted by Raimbault *et, al.* [5], Axelsson *et, al.* [6], William *et, al.* [7], Kawaii *et, al.* [8], Zeilter and Hellbrück [9] and the others in the past decade. The results statistically show that the principal components or major factors that are associated with *Dynamics*, *Temporal change* and similar subjective indices such as *Vibrancy*, *Eventfulness* and *Activity*, explain the second to fourth largest portion of the total variance after those with *Pleasantness* and *Calm*, which denote the considerable status of the temporal aspect in the soundscape assessment.

However, the temporal factor of sounds is a system that is too complex for full and accurate description and for finding suitable physical indicators, compared with other aspects such as loudness, spectrum and power. First, temporal structure has abundant details that cannot be covered by single indicator or variable. Second, environmental sounds contain different types of stochastic processes, and are much more indefinite than some simple signals (e.g., pure tone and white noise). Third, the fluctuations of different physical variables have a speed, magnitude and pattern of their own, and it is problematic to combine them in an appropriate way. Therefore, it is worthwhile to introduce temporal analysis in soundscape research and correspondingly construct a complete system for soundscape quality assessment.

© S. Hirzel Verlag · EAA

In 2003, Coensel *et, al.* [10] analysed rural and urban soundscape recordings, and found $1/f$ behaviours, a linear dependence on a log-log scale, in the A-weighted sound pressure, Zwicker loudness and instantaneous zero transitions (pitch) fluctuation spectra, which implies a critical balance of environmental sounds between predictability and chaos as that of classical music pieces. As they explained, this phenomenon originates from the universal complexity in physical environments. According to the theory of complex systems, $1/f$ related factors are available in the sonic environment, which is formed by ecosystems, natural and human social systems, etc. There have been few recent studies on related aspects, however, and they have only focused on individual sounds, such as wind [11, 12], human agglomeration [13] and traffic noise [14, 15].

Botteldooren *et, al.* [16] proposed a single indicator music-likeness to characterize the temporal structure of soundscape by quantitatively measuring the linear-fitting slope and deviation of the fluctuation spectrum. The result, which has $1/f$ behaviour is considered to be music-like, as indicated by fuzzy set membership functions,

$$ML = \mu_A(\alpha) \cdot \mu_E(\varepsilon), \tag{1}$$

where $\mu_A(\alpha)$ and $\mu_E(\varepsilon) \in (0, 1)$ are fuzzy sets of slope and deviation, respectively, and $ML \in (0, 1)$. An *ML* with a value that is approximately 1 indicates a high degree of music-likeness, while an *ML* that is approximately 0 indicates a low degree. Coensel's [17] preliminary research on the availability of *ML* used a field survey of passers-by to illustrate the feeling to the tempo in ten urban soundscapes. The results showed some correlation between *ML*1 (*ML* in the low frequency region) and the subject index *The distance from music-like*. However, there appeared to be large deviations of the data because of several factors including individual interest, sonic-environmental change and sample size.

Papers [18] and [19] proposed a multi-criteria assessment for quiet rural soundscapes that included *ML* and other five indicators: a pleasantness evaluation based on semantic differential method *SD-Pleasing*, non-fitting sounds, L_{A50}, noise event count N_{cn} (the number of sound events that exceed L_{A50} by 3 dBA) and $\log_{10}G$ (where G is the center of gravity of spectrum). This set of criteria was able to provide a guideline for soundscape design and creation, and was able to facilitate environmental monitoring and policy planning.

This paper reports on a study of the correlation between temporal parameters and an assessment of soundscape excerpts. In Section 2, the calculations of the fluctuation spectrum and the auto-correlation function (ACF) and parameters extracted from them are introduced, and soundscape excerpts are clustered on the basis of these parameters in Section 3. In Section 4, the arrangement of a laboratory listening test to assess the soundscape excerpts and construct their perceptual configuration is discussed. In Section 5, the correlation is analysed, first by PCA patterns of soundscape clusters distribution, and then by correlation coefficients and by stepwise multiple linear regression.

2. Parameter calculation

2.1. Samples of soundscape recordings

A total of 81 samples from soundscape recordings were gathered from two sources: a database of sound that was developed by Professor Guoqi Li in Beijing Union University [20] and the Freesound website [21]. The former is an accumulation of a huge amount of raw sound materials that were recorded by professional equipment, and the sound quality is therefore high enough for electro-acoustic system testing. The latter is a repository of Creative Commons licensed audio samples. The audio files that are upload by its users cover a wide range of subjects, including some elements of soundscape.

The samples were extracted from the recordings with a mono channel and duration of approximately 1 minute. There are roughly four aspects of the content of the samples, which can be expressed as nature (e.g., surf, river, thunder and wind), animals (e.g., bird, frog and cattle), human (e.g., shopping, social activity and playing games) and objects (e.g., billiard, fireworks and chimes).

2.2. Discrete sequences

Discrete sequences of short-time intensity and pitch were determined at a 100 Hz sampling frequency (interval of 10 ms). A relative A-weighted level L_{Aeq} (unknown absolute value) was employed to describe the intensity variation. The pitch indicator was defined as $\log_{10}P$, where P (in Hz) stands for *Correlogram-based Pitch Measures* in Slaney's MATLAB toolbox for auditory modelling work [22], which implements several popular auditory models for the MATLAB numerical programming environment. The pitch calculation procedure is now described, according to Pitch Detector proposed by Slaney and Lyon [23].

First, Lyon's Passive Cochlear Model, which is based on a transmission line model of the basilar membrane and is followed by several stages of adaptation, was employed to mimic the working of the auditory system that transmits sound pressure signals to auditory nerve firing. The model outputs an $N \times L$ matrix, where N is the number of cochlear channels and L is the length of the signal.

Second, multiple frames of a correlogram were computed, and each frame was stored as one row in a large array with a sampling frequency of 100 Hz. The result from the first step was input, and each image in the output correlogram is of size $N \times W$, where W is the length of the autocorrelation sequence and has a value of 256.

Third, on each frame of the correlogram, the pitch was estimated. The position of the largest peak in the frame was found inside the valid pitch range and considered to be an estimate of the pitch.

2.3. Fuzzy set memberships of the parameters extracted from the fluctuation spectra

The DC component in the discrete sequence that was mentioned in Section 2.2.1 was removed and then a Fourier

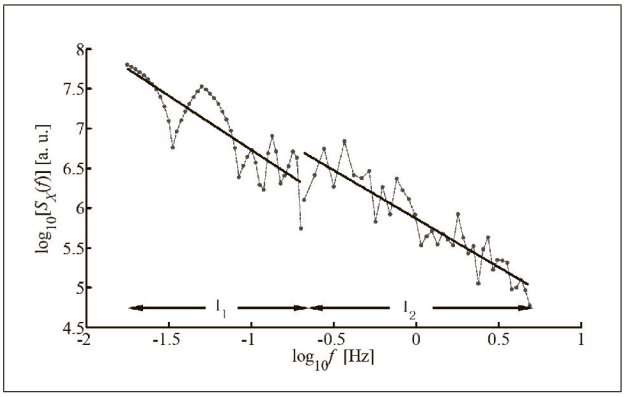

Figure 1. The patterns of the fluctuation spectrum processed to 12 points/octave (broken lines with dots) and their linear-fitting (thick lines) in the frequency region I_1 and I_2, respectively.

Transform $X(f)$ was carried out such that power spectrum could be derived as

$$S_X(f) = X(f)\,X^*(f), \tag{2}$$

where X stands for either of the variant (L_{Aeq} or $\log_{10}P$), and "*" denotes the complex conjugate of $X(f)$. The coordinates were transformed to log-log scale, and the spectra were then processed to 12 points per octave [17]. Interpolation must be used in the low frequency region I_1 =[0.017 Hz, 0.2 Hz] (time scale I_1' =[5 s, 60 s]) where the interval was larger than 1/12 Oct, while in the high frequency region I_2 =[0.2 Hz, 5 Hz] (time scale I_2' =[0.2 s, 5 s]), where the interval was small enough, averaging the energy of all of the points within each 1/12 Oct was used for smoothing. The cut-off point 5 s between I_1' and I_2' corresponds to the length of a phrase in musicology [17]. Finally, using linear fits for the spectra of intensity and pitch vibration in I_1 and I_2, the values of slopes α_{L1}, α_{L2}, α_{P1} and α_{P2} and the RMS deviations ε_{L1}, ε_{L2}, ε_{P1} and ε_{P2} were determined, where the subscripts L and P stand for L_{Aeq} and $\log_{10}P$, respectively, and 1 and 2 denote I_1 and I_2, respectively. An example of the fitting is shown in Figure 1. The eight parameters were finally transformed to fuzzy set memberships of music-likeness according to Botteldooren *et, al.* [16].

2.4. Parameters extracted from the autocorrelation function (ACF)

The ACF was first employed in acoustic research by Ando et, al [30] for preference tests of sound fields in concert halls. In the current study, however, the ACFs of discrete sequences of variables rather than the original signals were computed. The former represent the perceptible temporal fluctuations of the sounds, while the latter are perceived as other aspects such as pitch and timbre.

Initially, six parameters that were extracted from the ACF $\Phi(\tau)$ (or normalized ACF $\phi(\tau)$) were taken into consideration: $\phi_{\max}$ the amplitude of the maximum peak in $\phi(\tau)$, $\tau_{\max}$ the time delay of $\phi_{\max}$, $W_{\phi(0)/2}$ the width of attenuation from $\tau = 0$ with $\phi(0) = 1$ to the first time

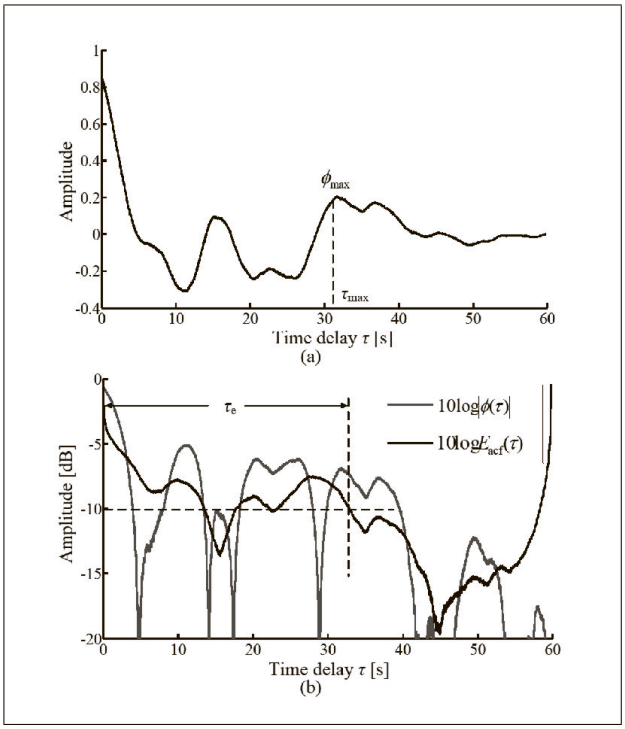

Figure 2. Definitions of four parameters extracted from the ACF: (a) $\phi_{\max}$ and $\tau_{\max}$, (b) τ_{e}.

lag where $\phi(\tau) = 0.5$, τ_{e} the effective duration of the ACF defined by a 10 dB delay of $\phi(\tau)$, the Envelope Magnitude Difference (*EMD*) and the Envelope Smoothness Measurement (*ESM*). The definitions of $\phi_{\max}$ and $\tau_{\max}$ are shown in Figure 2(a), while that of τ_{e} is shown in Figure 2(b). There are several ways to determine the value of τ_{e}, but in the current study, the ACF envelope was used,

$$E_{\mathrm{acf}}(\tau) = \sqrt{\Phi(\tau)^2 + \hat{\Phi}(\tau)^2}, \tag{3}$$

where $\hat{\Phi}(\tau)$ is the Hilbert transform of $\Phi(\tau)$. Therefore, τ_e can be defined as the duration of E_{acf} when attenuated by 10 dB. In addition, *EMD* and *ESM* (in the interval 0-τ' s; $\tau' < \tau$), as proposed by Liang Yan and Ke-an Chen [31], were also calculated in terms of E_{acf},

$$EMD = \frac{\sum_{i=1}^{\tau'\cdot f_s} |E_{\mathrm{acf}}(i+1) - E_{\mathrm{acf}}(i)|}{\tau' \cdot f_s}, \tag{4}$$

$$ESM = \frac{\tau' \cdot f_s \cdot \prod_{i=1}^{\tau'\cdot f_s} E_{\mathrm{acf}}(i)}{\sum_{i=1}^{\tau'\cdot f_s} E_{\mathrm{acf}}(i)}, \tag{5}$$

where f_s=100 Hz is the sample rate of the discrete sequences. Equations (4) and (5) imply that *EMD* quantifies the degree of fluctuation E_{acf}, while *ESM* can be used to measure the decay rate of E_{acf}.

It was observed from the resulting ACFs of $\log_{10}P$ (not shown here) that there were no obvious peaks in any of the samples, and it was difficult to obtain accurate values of $\phi_{\max}$ and $\tau_{\max}$. In addition, $W_{\phi(0)/2}$ was too small to be measured. Those parameters would make little sense and

Table I. Literatures on soundscape categorization on the basis of cluster algorithm.

Literatures	Cluster algorithm	Samples	Parameters	Numbers clusters
Coensel *et, al.* [24]	Artificial neural network	Rural and urban soundscapes, music,speech, etc	slopes and deviations of fluctuation spectrum	3
Coensel *et, al.* [25, 26]	Fuzzy ants Optimized by genetic algorithm	Soundscape recordings in urban parks	A-weighted level, spectrum and fluctuation spectrum	49
Gygi, Kidd and Watson [27]	Agglomerative hierarchical clustering	short (average 2.3s) environmental sounds	Subjective similarities	6
Rychtáriková and Vermeir [28]	Agglomerative hierarchical clustering	binaurally recordings in urban public places	Levels, sharpness, fluctuating strengths, roughness, etc	20
Jeon and Hong [29]	Hierarchical cluster analysis	Soundscape recordings in urban parks	perceptions of acoustic environments	3

thus should be abandoned. The remaining nine parameters were $\phi_{\max L}$, $\tau_{\max L}$, $W_{\phi(0)/2L}$, τ_{eL}, EMD_L, ESM_L, τ_{eP}, EMD_P and ESM_P, where the subscript L denotes L_{Aeq} and P denotes $\log_{10}P$ as in Section 2.2.2.

3. Soundscape categorization

The reason for clustering or classification, according to Schafer [4] and Truax [32], is to discover similarities between various soundscapes, to try to learn how different parameter combinations affect human's general impressions of sounds, and finally to improve acoustic designs. In the past few years, several studies have been conducted on soundscape clustering, as shown in Table I. The first three papers in the table are works of the Acoustic Group of Ghent University which preliminary brought temporal parameters to clustering. Coensel *et, al.* [24] is simple with small numbers of samples and parameters. The work of Coensel *et, al.* [25, 26] are much greater with over one thousand samples and three aspects of parameters, that is A-weighted level, spectrum and fluctuation spectrum. The resulting clusters were validated against visitorąś perceptual measurements of soundscape quality. Rychtáriková and Vermeir [28] used statistical acoustical parameters and mainly focused on psychophysical assessments of soundscapes of urban public places. They didn't verify the clustering results by subjective investigations, but just described the features of the resulting 20 clusters. Unlike above works, Gygi, Kidd and Watson [27] and Jeon and Hong [29] classified the sounds based on perceptions and identified the classification using acoustical parameters combinations.

Several indicators for soundscape dynamics have already been proposed based on early music research [10, 16, 17, 18, 19, 33] extracted from fluctuation spectrum, and have been found in soundscape categorization as shown in Table I. Nevertheless, the importance of temporal factor and related indicators is still a hypothesis which has not been confirmed by quantitative analysis. In this section, the soundscape excerpts were clustered in a manner similar to that in the above works, but the clusters were based only on the temporal parameters that were described in Section 2. The totally 17 parameters (eight from the fluctuation spectra and nine from ACFs) form a combination to cover the pattern of temporal structure and it is expected to find novel features of different types of sounds.

3.1. Bayesian Cluster Ensemble

There is little a priori knowledge regarding the temporal aspects of soundscapes, and thus a single cluster algorithm cannot guarantee the accuracy of a categorization. In this case cluster ensembles were available to provide a framework for combining multiple base clusterings of a dataset to generate a stable and robust consensus clustering. Bayesian Cluster Ensemble (BCE) [34], which solves the basic cluster ensemble approach using a Bayesian approach, was employed for the soundscape categorization based on temporal parameters because it outperforms other current algorithms such as cluster-based similarity partitioning algorithm (CSPA), hypergraph partitioning algorithm (HGPA) and meta-clustering algorithm (MCLA) [35], in terms of stability and accuracy. The general procedures of BCE are as follows:

First, K-means with different initializations were used to generate a set of base clusterings, and the default number of base clusters was set to 20.

Second, BCE was applied to the set of base clusterings to generate a consensus clustering by two steps, of which the first was a BCE model construction that was based on a base clustering matrix, and the second step was model parameter estimation.

BCE can solve the basic cluster ensemble approach by effectively maintaining a distribution over all possible consensus clusterings. It treats all base clustering results for each object as a feature vector with discrete feature values, and learns a mixed-membership model from the feature representation. One of the methods for learning a Bayesian cluster ensemble is Variational Inference, which estimates the model parameters and then mixed-membership. Expectation maximization (EM) algorithms

Table II. Samples of the soundscape excerpts in the nine resulting clusters. The underlined numbers indicate the samples selected for listening test in the next section.

Cluster	Samples	Features
1	3, 10, 22, 43, 46, 59, 74	none
2	19, 21, 26, 29, 31, 33, 35	small animals
3	13, 30, 37, 55, 58, 69	sound mixture
4	1, 5, 57	surf
5	2, 6, 7, 8, 9, 20, 61, 66, 67, 81	wind
6	4, 11, 15, 16, 17, 18, 23, 24, 28, 32, 34, 45, 47, 48, 49, 52, 60, 62, 63, 64, 65, 68, 71, 77, 79	water and public places
7	14	thunder
8	12, 25, 36, 38, 42, 44, 50, 51, 54, 56, 70, 72, 78	none
9	27, 39, 40, 41, 53, 73, 76, 80	livestock

are typically used for such parameter estimation problems by alternating between calculating the posterior over latent variables and updating the model parameters until convergence is achieved.

3.2. Results

Before clustering, the values of each of the 17 parameters with different units and scales were normalized to $(M - \langle M \rangle)/\sigma_M$, where M is a temporal parameter, σ_M is its SD value, and $\langle \bullet \rangle$ is the mean-value operator. There are two reasons for the normalization. First, the data of parameters inputted into the clustering algorithms is required to be normalized. Second, the values of the parameters of different clusters need to show in the same coordinate in the following and normalization is beneficial to indicate of their distributions.

The parameter estimation of the base clustering matrix in BCE was performed using EM algorithms with an iterative procedure. The results of the clustering are shown in Figure 3, in which the values of the 17 parameters in the resulting nine clusters are normalized. The samples of the excerpts were numbered, and Table II shows the samples that were contained in each cluster.

The patterns shown in Figure 3 indicate that most of the clusters have one or two parameters with extreme values and small enough deviations, such as ε_{P1} of cluster 1, ε_{L2} of cluster 2, EMD_P and α_{P2} of cluster 9, etc. Cluster 7 with only one sample (14) is even significantly different from other samples in τ_{eP}, ε_{L2} and α_{P1}. In conclusion, the procedure of BCE successfully classify the samples into nine clusters in different positions of the multi-dimensional parameter space.

In the following, the general features of the resulting nine clusters are analysed qualitatively according to the

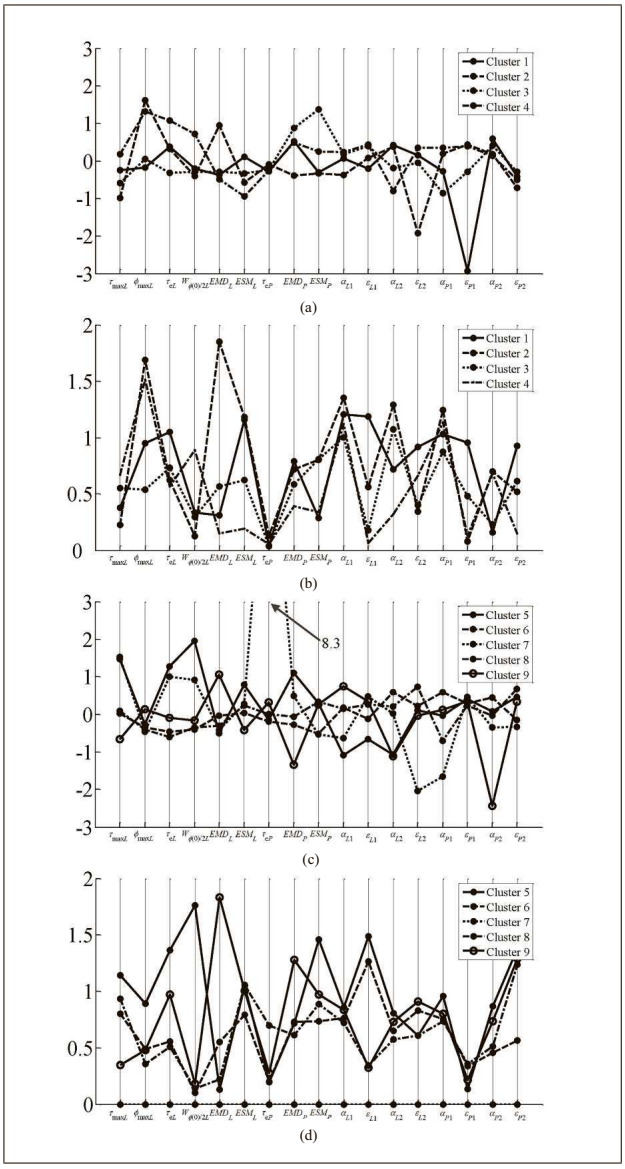

Figure 3. The patterns of averaged normalized values (a and c) and standard deviations (b and d) of the 17 parameters. (a) and (b) are patterns of cluster 1-4, while (c) and (d) are those of cluster 5-9.

results shown in Figure 3 and Table II. Cluster 1 contains seven samples that do not have distinguishing features but have very large ε_{L1}. Cluster 2 contains eight samples, most of which are sounds of small animals such as frogs (31, 33, 35), birds (21, 26) and crickets (29, 75). The samples sound short and sharp and have large $\phi_{\max L}$ and EMD_L. They also have large ε_{L2}. Cluster 3 contains six samples, including human footsteps (55, 58, 69), the rumblings of thunder (13), distant bird twitters, the babble of water (30) and the buzzing of bees (37). They are not intense, but have many sound sources that are mixed together, and have large ESM_P. Cluster 4 contains three samples, including surf (1, 5) and murmuring in a church (57); the $\phi_{\max L}$ and $W_{\phi(0)/2}$ are relatively large but ESM_L is small. Therefore, their intensity fluctuations are large but smooth. Cluster 5 contains 10 samples, which are primarily wind turbu-

Table III. The attributes and scales for the soundscape assessment which were translated from Chinese into English.

Attributes	Scales					Attributes
Pleasant	*very*	*a little*	*medium*	*a little*	*very*	*Unpleasant*
Evenful	*very*	*a little*	*medium*	*a little*	*very*	*UnEvenful*
Harmonic	*very*	*a little*	*medium*	*a little*	*very*	*Inharmonic*
Exciting	*very*	*a little*	*medium*	*a little*	*very*	*Boring*
Calm	*very*	*a little*	*medium*	*a little*	*very*	*Chaotic*
Fast	*very*	*a little*	*medium*	*a little*	*very*	*Slow*
Intensive	*very*	*a little*	*medium*	*a little*	*very*	*Gentle*
Static	*very*	*a little*	*medium*	*a little*	*very*	*Dynamic*
Simple	*very*	*a little*	*medium*	*a little*	*very*	*Complex*
Stressful	*very*	*a little*	*medium*	*a little*	*very*	*Relaxing*

lence, vegetation rustling in the breeze (6, 7, 8, 9) and slowly passing vehicles. Their $\tau_{\max L}$ and $W_{\phi(0)/2}$ are large, whereas α_{L1} and α_{L2} are approximately -2, which caused the sounds to be dull and boring. Cluster 6 is the largest, and contains 25 samples. Many of the samples in this cluster are of the same types, such as water and rain (11, 15, 16, 17, 18), and murmuring in the public places (45, 47, 48, 49, 52, 60, 77, 79). The 17 parameters generally have medium average values. Cluster 7 has only one sample, which is an excerpt of distant thunder (14), with a τ_{eP} of extreme value, large $\tau_{\max L}$ and ε_{L2}, and an α_{P1} of approximately -2. Cluster 8 contains 13 samples that do not have prominent feature, which is similar to cluster 6. Cluster 9 contains eight samples, including livestock, e.g., cattle and sheep (39, 40, 41), and chimes (73, 80); they have large EMD_L, small EMD_P and an α_{P2} that is approximately -2.

The result apparently indicates that these group of temporal parameters can explain some aspects of similarities among the samples which were properly classified into different clusters. And the procedure of BCE improved the accurateness mathematically. Except for cluster 1 and 8, the clusters generally contain a certain number of samples with similar sources. some of the clusters, such as cluster 6, contain more than one groups of samples with semeiological difference, because other factors were not considered.

4. Listening test

For the purpose of studying the correlation between temporal parameters and the soundscape assessment, a laboratory listening test was conducted to construct a pattern of auditory impressions on the soundscape excerpts.

4.1. Signals

Several representative samples (the underlined numbers in Table II) were selected in the nine clusters for the listening test. In all, 39 samples were used. They contained different types of sounds, and the values of their parameters reflected the features of their own clusters.

The audio files were played using Windows Media Player on a personal computer whose sound card was connected to a headphone amplifier (Brüel&Kjær ZE0769F). The samples were monaurally reproduced in headphones (Sennheiser HD580). The convolutions of the signals and the inverse impulse response of the headphones were computed for the frequency calibration. The signals were then reproduced onto a HEAD dummy head and the binaural pressure levels were measured using a Symphony Acoustic Analyzer. The L_{Aeq} of each signal was fixed to 65 dBA by adjusting the amplitude of the signal and the volume of the amplifier.

4.2. Subjects

There were 50 subjects (27 males, and 23 females; ranging from 19-43 years of age), and the subjects included university students and academics, and people engaged in other occupations, such as engineering, advertising media and publishing. A substantial proportion of the subjects had experience with psychoacoustic detection or other similar tests.

4.3. Assessments

Based on a rating scale method, a set of 10 unidirectional attribute scales was created for assessing the 39 soundscape excerpts by referring to the works in [5] to [9]. These attributes were expressed as pairs of antonyms in Chinese with five scales that modified the attributes by the words *very*, *a little* and *medium*. The attributes and scales for the soundscape assessment were translated from Chinese into English as shown in Table III.

The rating scale method was employed for statistical analysis. The cumulative frequencies of the 50 subjects' assessments to each of the attributes of each sample were transferred to scale values according to the Thurstone Model, which assumes that the preference probability to an object follows a Gaussian distribution. The 10 attribute scales of the 39 samples were further combined into four components using a PCA.

4.4. Procedures

The listening test was conducted in a semi-anechoic chamber, whose background noise can be neglected compared with the 65 dBA level of the signals. The amplifier could be connected to four pairs of headphones for testing at most four subjects at the same time. The samples were reproduced randomly, that is they were not sorted by the

result of the clustering or the types of sounds. The subjects were required to begin to click the attribute scales after a half duration of each signal was played when they had a general auditory impression of the sound. They were required to remove the headphones every twenty minutes to take a break for ten minutes. Each subject required approximately a half hour to finish the test.

4.5. Results

The results of the PCA for the soundscape excerpt assessment are shown in Table IV and Table V. Table IV indicates that the first to the fourth principle component (PC1 to PC4) explain 56.0%, 26.4%, 8.4% and 4.8% of the total variance, respectively, and their cumulative contribution reaches 95.6%.

The eigenvectors that are shown in Table V indicate the loadings of the 10 attribute scales to PC1 to PC4. The loadings of most of the attributes in PC1 are large except for *Evenful-Unevenful* and *Exciting-Boring*, The first four largest attributes are *Calm-Chaotic*, *Stressful-Relaxing*, *Harmonic-Inharmonic* and *Simple-Complex*. On the contrary, PC2 is associated with *Evenful-Unevenful* and *Exciting-Boring*. The greatest loading of PC3 is *Fast-Slow*, followed by *Simple-Complex* and *Intensive-Gentle*. PC4 is associated with *Calm-Chaotic*, *Static-Dynamic* and *Evenful-Unevenful*.

The result is basically consistent with those of previous investigations [5, 6, 7, 8, 9] and just has slightly different features for distinctions of procedures, samples and attributes selections, backgrounds of the subjects and so on. The factor or component which contributes most to the total variance in previous results was labeled as *Pleasantness*, *Relaxation* and *Calm*, and they are also found to have large loadings in PC1 as shown in Table V. PC2 are almost identical to previous results which was usually labeled as *Evenfulness* and *Vibrancy*. *Fast-Slow* has much greater loadings (0.72) than other attributes in PC3 implying that it represents rhythm of the sounds. And PC4 represents dynamics for it first two loadings *Calm-Chaotic* (0.56) and *Static-Dynamic* −0.54. PC3 and PC4 make relatively small contribution (13.2%) to the total variance compared to PC1 and PC2, but they must not be disregarded for they may explain the details in the perception. Hence, PC1 to PC4 will be examined in the following discussion.

Table IV. Contributions and cumulative contributions of PC1 to PC4.

PC	Contribution/(%)	cumulative Contribution/(%)
1	56.0	56.0
2	26.4	82.4
3	8.4	90.8
4	4.8	95.6

Table V. Eigenvectors or loadings of PC1 to PC4, where the underlined numbers indicate the first three greatest loadings of the PCs (the numbers are four in PC1 because there were same numbers), and the attributes in the first column are translations from Chinese into English.

Attributes	PC1	PC2	PC3	PC4
Pleasant-Unpleasant	0.32	0.25	0.28	−0.01
Evenful-Unevenful	−0.08	0.46	−0.05	0.39
Harmonic-Inharmonic	0.36	0.25	0.13	0.05
Exciting-Boring	−0.03	0.55	0.22	−0.26
Calm-Chaotic	0.48	−0.16	0.25	0.56
Fast-Slow	−0.31	0.01	0.72	−0.23
Intensive-Gentle	−0.32	−0.11	0.35	0.17
Static-Dynamic	0.25	−0.24	−0.11	−0.54
Simple-Complex	0.36	−0.39	0.36	−0.09
Stressful-Relaxing	−0.37	−0.35	0.04	0.29

5. Correlation analysis

5.1. Sample distribution in PC plots

In the first step of the analysis, a pattern of the sample distribution was shown to qualitatively describe perceptual features of clusters with different parameter combinations. Figures 4(a) and (b) display the component scores of the 39 samples in PC1-PC2 and PC3-PC4, respectively. These two plots indicate some concentrations of samples of the same clusters, such as cluster 2 and cluster 9. However, some other clusters have exceptional samples that are located far from the zones of their own clusters. For example, Sample 11 and 15 have significant differences in the scores in PC1 and PC3 from other samples that are in cluster 6. The deviations of the scores in the four PCs in the same cluster are generally different. For example, the samples of cluster 5 are concentrated in PC2 but are dispersed in PC1, while the concentrations of cluster 8 are reversed.

The results that are shown in Figure 4 indicate that clustering on the basis of temporal parameters can correspond to the soundscape assessment data to some extent, However, as discussed above, the pattern is not perfect for three possible reasons.

First, many other objective and subjective factors that also have an effect on perception to sounds, were not included in the clustering, especially undetectable information regarding social, cultural, semantic and warning aspects and psychological preferences to some elements. For example, most of the subjects preferred samples 11 and 15 to other samples in cluster 6 according to the scores of PC1, because the former samples are natural soundscapes

(rain and river flow), while the latter are noisy voices in public places.

Second, none of the current clustering algorithms can achieve an accurate categorization. K-means with different initializations must generate various sets of base clusterings; thus the consensus clustering would not be unique in spite of the step of cluster ensembles.

Third, the parameters were not weighted in the clustering algorithms, but were equally important. However, their effects on perception may be distinctly different, and it is difficult to estimate which role each parameter is playing when there is a lack of a priori knowledge. Evidence that qualitatively reveals this problem is shown in Figure 4b. Cluster 7 contains only one sample labelled 14 that has an extremely large τ_{eP} as shown in Figure 3c. Nevertheless, it did not correspondingly induce an extreme assessment, and the sample is only located in the upper left of the PC3-PC4 (their cumulative contribution is only 13.2%) coordinate and is separate from the other samples.

5.2. Relationships between the PCs and temporal parameters

5.2.1. Correlation calculation and discussion

In this section, a quantitative analysis to further study the relationships between the scores of the four PCs and 17 temporal parameters is reported.

Pearson's (r_p) and Spearman's (r_s) coefficients are given in Tables VI and VII. The former is a measure of the linear correlation between two variables, and the latter is a nonparametric rank statistic that is used to measure the monotonicity between two variables without relation to the data distribution.

There are several strong correlations between the nine parameters extracted from ACFs, for example those between $\tau_{\max L}$ and $W_{\phi(0)/2L}$ ($r_p = 0.50$ ($p \leq 0.001$), $r_s = 0.47$ ($p \leq 0.01$)), ESM_L ($r_p = 0.51$ ($p \leq 0.001$), $r_s = 0.42$ ($p \leq 0.01$)), EMD_P ($r_p = 0.45$ ($p \leq 0.01$), $r_s = 0.46$ ($p \leq 0.01$)), etc. It must be redundant for this group of parameters to describe ACF to a certain extent, and number of ACF parameters needs to reduce further. The coefficients between the eight parameters extracted from fluctuation spectra are relatively small, except those between α_{L1} and α_{P1} ($r_p = 0.38$ ($p \leq 0.05$), $r_s = 0.43$ ($p \leq 0.01$)), and between α_{L2} and α_{P2} ($r_p = 0.54$ ($p \leq 0.001$), $r_s = 0.46$ ($p \leq 0.01$)), which implies that the fluctuations of intensity and pitch of soundscape match with each other in different time scales. It can also be observed that parameters from ACFs and fluctuation spectra are generally weak. ACF and fluctuation spectrum are time and frequency domain indication of the sequence, respectively, related to each other by Wiener-Khintchine theorem. However, they provide considerably different information about temporal fluctuation according to the results in Table VI and VII.

The crosswise comparison between the Pearson's and Spearman's coefficients shows that most of their values r_p and r_s are consistent with each other. But there are also a few exceptions. Low r_p but high r_s such as τ_{eP} and EMD_L ($r_p = -0.02$ ($p > 0.05$), $r_s = 0.47$ ($p \leq 0.01$)) indicate

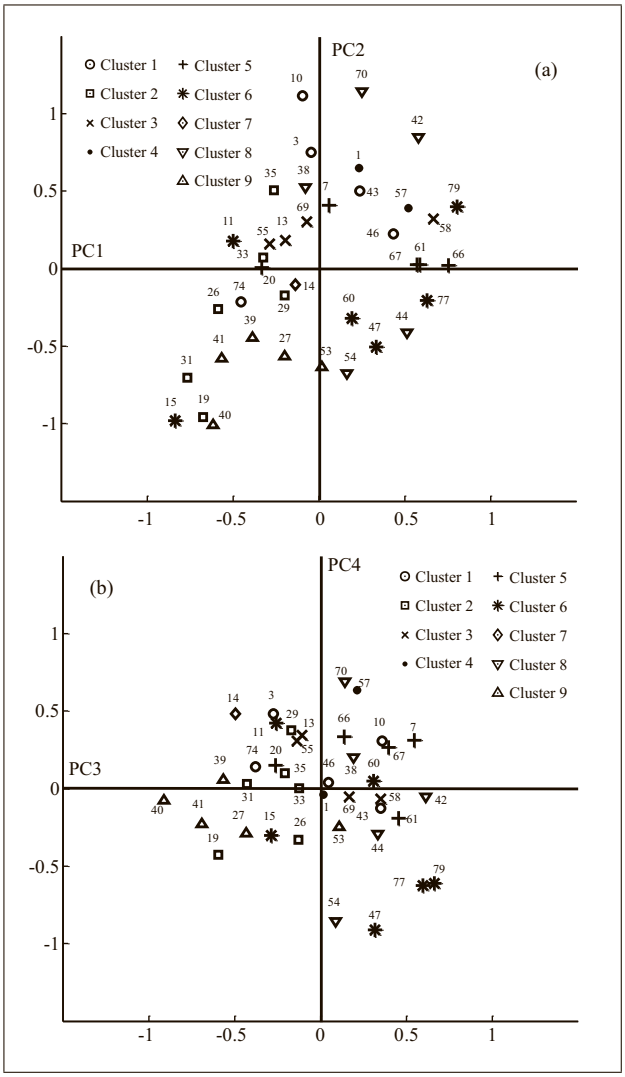

Figure 4. The component scores of the 39 samples in (a) PC1-PC2 and (b) PC3-PC4.

the pair are just monotonously but not linearly dependent, originated from sample 14 with extremely large τ_{eP}.

PC1 is most highly correlated to EMD_L ($r_p = -0.46$ ($p \leq 0.01$), $r_s = -0.40$ ($p \leq 0.05$)), EMD_P ($r_p = 0.33$ ($p \leq 0.05$), $r_s = 0.31$ ($p > 0.05$)) and ε_{L2} ($r_p = 0.45$ ($p \leq 0.01$), $r_s = 0.42$ ($p \leq 0.01$)). EMD_L and EMD_P are negatively correlated to each other, thus they must be collinear when explaining PC1.

PC2 is most highly correlated to EMD_L ($r_p = -0.45$ ($p \leq 0.01$), $r_s = -0.56$ ($p \leq 0.001$)), EMD_P ($r_p = 0.46$ ($p \leq 0.01$), $r_s = 0.48$ ($p \leq 0.01$)), α_{P1} ($r_p = -0.41$ ($p \leq 0.01$), $r_s = -0.42$ ($p \leq 0.01$)) and α_{P2} ($r_p = 0.47$ ($p \leq 0.01$), $r_s = 0.37$ ($p \leq 0.05$)). Moreover, PC2 are also dependent on α_{P1} ($r_p = -0.41$ ($p \leq 0.01$), $r_s = -0.42$ ($p \leq 0.01$)) and α_{P2} ($r_p = 0.47$ ($p \leq 0.01$), $r_s = 0.37$ ($p \leq 0.05$)), the slopes of pitch fluctuation spectrum at large and small time scale.

The effects of EMD_L ($r_p = -0.32$ ($p \leq 0.05$), $r_s = -0.30$ ($p > 0.05$)) and EMD_P ($r_p = 0.33$, ($p \leq 0.05$), $r_s = 0.29$ ($p > 0.05$)) on PC3 are also observable, but the significance levels are relatively low and the effects are

Table VI. Pearson's (upper) and Spearman's (lower) coefficients between the values of the parameters (on the basis of parameter calculation of 81 soundscape excerpts) and the scores of PC1 PC4 (on the basis of the data of assessments to 39 soundscape excerpts) (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$).

	$\tau_{\max L}$	$\phi_{\max L}$	τ_{eL}	$W_{\phi(0)/2L}$	EMD_L	ESM_L	τ_{eP}	EMD_P	ESM_P
$\phi_{\max L}$	−0.34*								
	−0.39*								
τ_{eL}	0.41**	0.33*							
	0.26	0.49**							
$W_{\phi(0)/2L}$	0.50***	0.10	0.52***						
	0.47**	0.23	0.60***						
EMD_L	−0.37*	0.15	−0.10	−0.28					
	−0.44**	0.33*	−0.14	−0.26					
ESM_L	0.51***	−0.16	0.26	0.17	0.03				
	0.42**	−0.14	0.16	0.06	−0.01				
τ_{eP}	0.26	−0.06	0.16	0.23	−0.02	0.02			
	−0.17	0.22	0.11	0.12	0.47**	−0.14			
EMD_P	0.45**	−0.14	0.24	0.39*	−0.44**	0.10	−0.01		
	0.46**	−0.10	0.21	0.31	−0.58***	0.07	−0.40*		
ESM_P	−0.13	0.02	−0.12	−0.14	−0.06	−0.15	−0.07	0.1	
	−0.06	0.08	−0.08	−0.01	−0.03	−0.18	0.18	0.13	
α_{L1}	−0.20	−0.19	−0.24	−0.30	−0.12	−0.46**	−0.06	−0.23	0.10
	−0.09	−0.21	−0.32*	−0.08	0.04	−0.34*	0.13	−0.30	0.00
ε_{L1}	−0.11	−0.26	−0.34*	0.12	0.01	−0.22	0.05	−0.19	0.11
	−0.05	−0.12	−0.15	0.08	−0.16	−0.04	0.10	0.17	0.16
α_{L2}	−0.18	−0.08	−0.43**	−0.50***	0.11	0.12	−0.04	−0.08	0.05
	−0.21	−0.30	−0.42**	−0.79***	0.09	0.18	−0.18	−0.18	0.00
ε_{L2}	0.26	−0.36*	−0.10	0.18	−0.33*	0.06	−0.31	−0.01	0.17
	0.29	−0.18	−0.07	0.31*	−0.12	0.06	−0.13	−0.08	0.19
α_{P1}	−0.13	−0.04	−0.15	−0.28	0.27	−0.14	−0.21	−0.38*	−0.23
	−0.12	−0.06	−0.21	−0.19	0.26	−0.15	0.26	−0.37*	−0.22
ε_{P1}	0.05	0.13	−0.07	0.14	0.22	0.10	0.13	−0.29	−0.01
	−0.13	0.22	−0.02	0.01	0.26	−0.07	0.35*	−0.38*	−0.29
α_{P2}	0.08	−0.13	−0.17	−0.02	−0.19	0.15	−0.17	0.45**	−0.02
	−0.04	−0.12	−0.04	−0.30	−0.37*	0.24	−0.78***	0.16	−0.27
ε_{P2}	0.46**	−0.11	0.35*	0.42**	−0.20	0.06	−0.01	0.04	0.11
	0.39*	−0.03	0.24	0.39*	−0.21	−0.14	0.14	0.06	0.29
PC1	0.24	−0.13	0.06	0.20	−0.46**	0.17	−0.08	0.33*	0.28
	0.28	−0.16	0.00	0.06	−0.40*	0.15	−0.03	0.31	0.34*
PC2	0.20	0.05	0.13	0.13	−0.45**	0.02	−0.11	0.46**	−0.03
	0.21	−0.03	0.13	0.00	−0.56***	0.00	−0.49***	0.48**	0.00
PC3	0.24	−0.23	−0.05	0.08	−0.32*	0.34*	−0.27	0.33*	0.16
	0.29	−0.28	−0.09	−0.09	−0.30	0.34*	−0.20	0.29	0.22
PC4	0.23	0.22	0.31	0.28	−0.11	−0.21	0.18	0.39*	−0.07
	0.23	0.25	0.33*	0.19	−0.30	−0.21	−0.21	0.43**	−0.04

then weak. Two other important parameters are ε_{L2} ($r_p = 0.43$ ($p \leq 0.01$), $r_s = 0.42$ ($p \leq 0.01$)) and ESM_L ($r_p = 0.34$ ($p \leq 0.05$), $r_s = 0.34$ ($p \leq 0.05$)). Linear correlation of α_{P1} is also significant ($r_p = 0.42$ ($p \leq 0.01$), $r_s = 0.10$ ($p > 0.05$)).

PC4 is only weakly correlated to EMD_P ($r_p = 0.39$ ($p \leq 0.05$), $r_s = 0.43$ ($p \leq 0.01$)) and α_{P1} ($r_p = -0.32$ ($p \leq 0.05$), $r_s = -0.34$ ($p \leq 0.05$)). Most of the parameters have coefficients of 0.20-0.30, but none of them is significant.

In general, not many parameters can significantly correlated to the PCs, and thus reflect the effects of temporal factor on soundscape perception. Most of the correlations are weak, and may even be negligible (e.g. τ_{eL}, $W_{\phi(0)/2L}$, α_{L1} and ε_{L1}). On the contrary, there are also three to four parameters with largest coefficients. EMD_L and EMD_P

Table VII. Pearson's (upper) and Spearman's (lower) coefficients between the values of the parameters (on the basis of parameter calculation of 81 soundscape excerpts) and the scores of PC1 PC4 (on the basis of the data of assessments to 39 soundscape excerpts) (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$) (continue).

	α_{L1}	ε_{L1}	α_{L2}	ε_{L2}	α_{P1}	ε_{P1}	α_{P2}	ε_{P2}
ε_{L1}	0.07							
	−0.20							
α_{L2}	−0.25	0.00						
ε_{L2}	0.19	0.24	−0.24					
	0.07	0.01	−0.28					
α_{P1}	0.38*	0.13	−0.03	−0.07				
	0.43**	0.09	−0.03	−0.09				
ε_{P1}	−0.08	0.30	−0.15	−0.03	0.29			
	−0.06	0.09	−0.05	−0.05	0.37*			
α_{P2}	−0.31*	0.03	0.54***	−0.08	−0.30	−0.27		
	−0.32*	−0.00	0.46**	−0.14	−0.30	−0.14		
ε_{P2}	−0.15	0.07	−0.24	0.30	−0.09	0.23	−0.14	
	−0.08	0.06	−0.24	0.32	−0.09	0.26	−0.23	
PC1	−0.00	−0.02	0.10	0.45**	−0.13	−0.03	0.19	0.14
	−0.08	0.03	0.08	0.42**	−0.16	−0.18	−0.06	0.24
PC2	−0.20	−0.16	0.14	0.22	−0.41**	−0.28	0.47**	0.08
	−0.24	−0.15	0.09	0.21	−0.42**	−0.36*	0.37*	0.15
PC3	−0.12	0.07	0.27	0.43**	−0.13	0.02	0.42**	0.13
	−0.12	0.06	0.17	0.42**	−0.12	−0.22	0.10	0.18
PC4	−0.22	−0.05	−0.06	−0.17	−0.32*	−0.18	0.10	0.20
	−0.29	−0.00	−0.02	−0.14	−0.34*	−0.13	0.14	0.27

are moderately correlated to PC1 and PC2 which explain 82.4% of the total variance. Samples with negative PC1 and PC2 scores generally have large EMD_L and small EMD_P (e.g. sample 31 the frogs and sample 40 and 41 the sheep), while those with positive PC1 and PC2 scores have small EMD_L and large EMD_P (e.g. sample 66 the street soundscape with traffic noise and sample 79 the murmur in a club). It seems that the subjects generally felt calm, relaxing and exciting to the animals but chaotic, stressful and boring to the elements of social life not only because of the meaning and c ontent of the sources, but also of the magnitude of temporal change. The strong correlations of ε_{L2} to PC1 and PC3 are remarkable. This results can be interpreted as the importance of intensity fluctuation of small time scale to perception, but it is difficult to understand the deviation of linear fitting represents which aspect of temporal structure. Finally, it is worthwhile to note that most of the pitch parameters, such as τ_{eP} ($r_p = -0.11$ ($p > 0.05$), $r_s = -0.49$ ($p \leq 0.01$)), α_{P1} ($r_p = -0.41$ ($p \leq 0.01$), $r_s = -0.42$ ($p \leq 0.01$)) and α_{P2} ($r_p = 0.47$ ($p \leq 0.01$), $r_s = 0.37$ ($p \leq 0.05$)) (α_{P2} is also highly linearly correlated to PC3), are significantly correlated to PC2, reflecting the effect of pitch fluctuation on this part of assessment.

5.2.2. Stepwise multiple linear regression

Four stepwise multiple linear regression models were constructed to determine the best and the simplest parameter combinations to explain PC1-PC4. Significance test is conducted when each new parameter is incorporated to the linear regression equation, and insignificant parameters should be rejected. The incorporation-rejection procedure is iterated until no significant parameter can be incorporated and no insignificant parameter can be rejected. The results are shown in Table VIII, where R^2 and R^2_{adj} are multiple correlation coefficient and adjusted multiple correlation coefficient, respectively, indicating the percentage of the final linear regression equation explaining the total variance of the PCs.

PC1 and PC3 are predominated by EMD_P and ε_{L2} in their regression models, similar to the results in Table VI and VII. The total variance of PC1, however, cannot be majorly explained by this parameter combination ($R^2_{\text{adj}} = 0.38$). Whereas that of PC3 is better explained ($R^2_{\text{adj}} = 0.51$) by adding α_{P2} ($\beta_j = 0.20$, $t = 2.21$ ($p \leq 0.001$)) to the model. Probably because of the redundancy among the individual parameters, the best multiple regression solutions often did not include the best individual parameters, such as the case of PC2 and PC4. The strongest parameters for PC2 Table VI, such as EMD_L, EMD_P, α_{P1} and α_{P2}, appear weaker in the regression model, and there is almost no dominating parameter. On the contrary, those insignificant for PC4 Table VI, such as $\tau_{\max L}$, $\phi_{\max L}$ and ESM_L, contribute most to the total variance in the regression model.

The results in Table VII indicate that the temporal parameters can explain the variance of PC scores to a certain extent, but many of the parameters added to the models are

Table VIII. Results of stepwise multiple linear regression analysis of PC1-PC4 (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.005$). β_j: Standardized multiple regression coefficient.

Model fit, R^2 (R^2_{adj})	F-change	Param.	β_j	t-value
PC1:				
0.54 (0.38)	0.30***	$\tau_{\max L}$	−0.23	−2.03
		τ_{eL}	0.13	1.71
		EMD_L	−0.26	−1.82
		ESM_L	0.10	1.18
		τ_{eP}	0.08	1.43
		EMD_P	0.23	2.69**
		α_{L2}	0.17	2.41*
		ε_{L2}	0.29	3.80***
		α_{P1}	0.09	1.30
		ε_{P1}	0.08	1.32
PC2:				
0.56 (0.42)	4.04***	$\phi_{\max L}$	0.16	2.33*
		$W_{\phi(0)/2L}$	−0.28	−1.94
		EMD_L	−0.18	−1.13
		EMD_P	0.22	2.14*
		ESM_P	0.16	1.99
		ε_{L2}	0.18	2.31*
		α_{P1}	−0.14	−1.72
		α_{P2}	0.12	1.79
		ε_{P2}	0.08	1.03
PC3:				
0.62 (0.51)	6.00***	$\tau_{\max L}$	−0.14	−1.84
		τ_{eL}	0.09	1.48
		ESM_L	0.15	2.21*
		EMD_P	0.22	3.89***
		α_{L1}	0.08	1.41
		α_{L2}	0.20	3.66***
		ε_{L2}	0.23	4.75***
		ε_{P1}	0.10	2.30*
PC4:				
0.50 (0.37)	3.75***	$\tau_{\max L}$	0.23	2.80**
		$\phi_{\max L}$	0.11	2.31*
		EMD_L	0.17	1.55
		ESM_L	−0.23	−3.36**
		EMD_P	0.10	1.57
		ESM_P	−0.07	−1.39
		α_{L2}	0.04	0.96
		α_{P1}	−0.12	−2.19*

weakly significant, and the R^2_{adj} are generally not larger than 50% except that of PC3. R^2_{adj} is called the adjusted multiple correlation coefficient which decreases when an insignificant parameter is added to the model, and it would be more preferable than R^2, which increases monotonically when more parameters are added. In addition to other factors that were involved in the soundscape perception, this result probably occurred because these temporal parameters do not linearly or multiple-linearly correspond to the assessment. The Pearson's coefficient in Table VI and VII and the stepwise multiple linear regression in Table VIII measure linear relations between variables, whereas the Spearman's coefficient measures monotonic relations. However, that may no t the case here. In our opinions, there could be two possible patterns. One is that the parameters without significant levels have nonlinear effects on the perception. The other is that the effects of these parameters are not independent but interact with one another, and different parameter combinations generate various impressions on the sounds, as the pattern of Figure 4.

6. Conclusions

In summary, underlying information about the temporal structure of environmental sounds was preliminarily discovered in this work, which provides a novel idea for the study of soundscapes and the sonic environment. The categorization and labeling of samples of soundscape recordings in Section 3 helps us find the similarities and distinctions between sound sources in the temporal aspect. The clustering result shows that several temporal parameters extracted from the recordings can describe source radiation features, and thus most of the samples with similar sources were classified into the same clusters. A listening test was conducted based on a rating scale method in Section 4 in order to verify the validity of the clustering. Totally 39 samples selected in the nine clusters were reproduced in headphones, and four PCs were extracted from the assessments of 50 subjects. The relationships between the PCs and temporal parameters was analysed in Section 5. Coefficients between the parameters and the PCs were calculated and stepwise multiple linear regression models were also given in this section. Although correlation calculation cannot completely explain the pattern, features of different clusters and significance of some parameters were found statistically, which indicates the necessity of introducing temporal aspect in soundscape research. Linear Regression models can just explain no more than 50% of the total variance. This implies that the effect of the temporal factor on soundscape assessment must be a complex system, there may be nonlinear effects on the perception, and these parameters may not be independent but interact with one another. Other novel methods for describe and evaluate tempo of soundscape are expected in the future.

Acknowledgements

The authors would like to thank Professor Guoqi Li in Beijing Union University for providing audio files of his database of sound. The authors would also like to thank the subjects who participated in the listening test.

References

[1] A. L. Brown: Soundscapes and environmental noise management. Noise Control Eng. J. **58** (2010) 493–500.

[2] S. Viollon, C. Lavandier: Multidimensional assessment of the acoustic quality of urban environments. Inter-Noise 2000, Nice, France, 2000.

[3] J. D. Guillén, I. L. Barrio: Importance of personal, attitudinal and contextual variables in the assessment of pleasantness of the urban sound environment. ICA 2007, Madrid, Spain, 2007.

[4] R. M. Schafer: Our sonic environment and the soundscape: the tuning of the world. Destiny Books, Rochester, Vermont, USA, 1994.

[5] M. Raimbault, C. Lavandier, M. Bérengier: Ambient sound assessment of urban environments: field studies in two french cities. Applied Acoustics. **64** (2003) 1241–1256.

[6] Ö. Axelsson, B. Berglund, M. E. Nilsson: Soundscape assessment. J. Acoust. Soc. Am. **117** (2005) 2591–2592.

[7] W. J. Davies, M. D. Adams, N. S. Bruce, *et al.*: Perception of soundscapes: an interdisciplinary approach. Applied Soundscapes. **74 (2)** (2013) 224–231.

[8] K. Kawai, T. Kojima, K. Hirate, M. Yasuoka: Personal evaluation structure of environmental sounds: experiments of subjective evaluation using subjects' own terms. J. Sound Vib. **227** (2004) 523–533.

[9] A. Zeitler, J. Hellbrück: Semantic attributes of environmental sounds and their correlations with psychoacoustic magnitudes. ICA 2001, Rome, Italy, 2001.

[10] B. de Coensel, D. Botteldooren, T. de Muer: 1/f noise in rural and urban soundscapes. Acta Acustica United With Acustica. **89** (2003) 287–295.

[11] S. Morgan, R. Raspet: Investigation of the mechanisms of low-frequency wind noise generation outdoors. J. Acoust. Soc. Am. **92** (1992) 1180–1183.

[12] V. E. Ostashev: Acoustics in moving inhomogeneous media. E&FN Spon, London, 1997.

[13] H. V. Ribeiro, R. T. de Souza, E. K. Lenzi, R. S. Mendes, L. R. Evangelista: The soundscape dynamics of human agglomeration. New Journal of Physics. **13 (2)** (2011) 1–8.

[14] X. Zhang, G. Hu: 1/f noise in a two lane highway traffic model. Phys. Rev. E. **52** (1995) 4664–4668.

[15] D. Chowdhury, A. Schadschneider: Self-organization of traffic jams in cities: Effects of stochastic dynamics and signal periods. Phys. Rev. E. **59 Part A** (1999) 1311–1314.

[16] D. Botteldooren, B. de Coensel, T. de Muer: The temporal structure of the urban soundscape. CFA/DAGA'04, Strasbourg, 2004.

[17] B. de Coensel: Introducing the temporal aspect in environmental soundscape research. Dissertation. Ghent University, 2007.

[18] B. de Coensel, D. Botteldooren: The quiet rural soundscape and how to characterize it. Acta Acustica United With Acustica. **92** (2006) 887–897.

[19] D. Botteldooren, B. de Coensel: Quality labels for the quiet rural soundscape. Inter-Noise 2006, Honolulu, Hawaii, USA, 2006.

[20] G. Q. Li: Soundscape study and soundscape design. Dissertation. Tsinghua University, 2004.

[21] http://www.freesound.org/.

[22] M. Slaney: Auditory toolbox: A matlab toolbox for auditory modeling work. Tech. Rept. Interval Research Corporation, 1998.

[23] M. Slaney, R. F. Lyon: A perceptual pitch detector. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Albuquerque, USA, 1990.

[24] B. de Coensel, D. Botteldooren, T. de Muer: Classification of soundscapes based on their dynamics. ICBEN 2003, Rotterdam, The Netherlands, 2003.

[25] B. de Coensel, D. Botteldooren, K. Debacq, *et al.*: Soundscape classifying ants. Inter-Noise 2007, Istanbul, Turkey, 2007.

[26] B. de Coensel, D. Botteldooren, K. Debacq, *et al.*: Clustering outdoor soundscapes using fuzzy ants. Congress on Evolutionary Computation, (IEEE World Congress on Computational Intelligence), Hong Kong, 2008.

[27] B. Gygi, G. R. Kidd, C. S. Watson: Similarity and categorization of environmental sounds. Perception & Psychophysics. **69(6)** (2007) 839–855.

[28] M. Rychtáriková, G. Vermeir: Soundscape categorization on the basis of objective acoustical parameters. Applied Acoustics. **74** (2013) 240–247.

[29] J. Y. Jeon, J. Y. Hong: Classification of urban park soundscapes through perceptions of the acoustical environments. Landscape and Urban Planning. **141** (2015) 100–111.

[30] Y. Ando, T. Okano, Y. Takezoe: The running autocorrelation function of different music signals relating to preferred temporal parameters of sound fields. J. Acoust. Soc. Am. **86 (2)** (1989) 644–649.

[31] Y. Liang, K. A. Chen: Auditory perceptual information extraction from acf physical factors of road-vehicle noise. ICCIS2010, Chengdu, China, 2010.

[32] B. Truax: Soundscape studies: an introduction to the world soundscape project. Numus West. **5** (1974) 37.

[33] R. F. Voss, J. Clarke: ąř1/f noiseąś in music: Music from 1/f noise. J. Acoust. Soc. Am. **63** (1978) 258–263.

[34] H. Wang, H. Shan, A. Banerjee: Bayesian cluster ensembles. Statistical Analysis and Data Mining. **4 (1)** (2011) 54–70.

[35] A. Strehl, J. Ghosh: Cluster ensembles-a knowledge reuse framework for combining multiple partitions. The Journal of Machine Learning Research. **3** (2002) 583–617.