[http://dx.doi.org/10.1016/j.jsv.2016.03.040]

Pitch features of environmental sounds

Ming Yang, Jian Kang
School of Architecture, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

* Corresponding author:
Email: j.kang@sheffield.ac.uk

Abstract

A number of soundscape studies have suggested the need for suitable parameters for soundscape measurement, in addition to the conventional acoustic parameters. This paper explores the applicability of pitch features that are often used in music analysis and their algorithms to environmental sounds. Based on the existing alternative pitch algorithms for simulating the perception of the auditory system and simplified algorithms for practical applications in the areas of music and speech, the applicable algorithms have been determined, considering common types of sound in everyday soundscapes. Considering a number of pitch parameters, including pitch value, pitch strength, and percentage of audible pitches over time, different pitch characteristics of various environmental sounds have been shown. Among the four sound categories, i.e. water, wind, birdsongs, and urban sounds, generally speaking, both water and wind sounds have low pitch values and pitch strengths; birdsongs have high pitch values and pitch strengths; and urban sounds have low pitch values and a relatively wide range of pitch strengths.

Keywords: environmental sound, pitch estimation, psychoacoustics, soundscape

2016 Journal of Sound and Vibration
Received 21 October 2015, **Revised** 16 March 2016, **Accepted** 28 March 2016,
Available online 14 April 2016

1 Introduction

Over the past fifteen years, the perception and evaluation of soundscape (referring to the total sound environment [1]) have been researched through numerous studies. It has been revealed that conventional acoustic parameters for noise measurement [2, 3], e.g. weighted sound pressure levels (SPLs), alone are not adequate for the measurement of soundscape [4]; more parameters are needed, which are more likely to be correlated with people's subjective evaluation of soundscape [5-7], such as comfort [4], pleasantness [8], annoyance [9], etc. For example, background noise level, standard deviation of short L_{Aeq} [4, 8], temporal structure [10, 11] and some psychoacoustic parameters [12, 13] have been used. In addition to these parameters, there is a recognized need to explore the possibility of additional parameters for soundscape measurement.

Since soundscape and music are closely related, in that music could be regarded as an imitation of environmental soundscapes or an ideal soundscape of the mind [1], musical features, particularly the psychoacoustic parameters that have previously been applied mainly in music perception, may also be applicable in soundscape research. In the fields of music psychology and psychoacoustics, the sensations of hearing are generally studied from four aspects, i.e. loudness, pitch, rhythm, and timbre. While loudness, timbre (including sharpness, tonality, roughness and fluctuation strength), and rhythm have been used to analyze the characteristics of soundscapes and environmental sounds [14-17], further study is required of the pitch aspect. Pitch corresponds to the sound's physical property of frequency, whereas loudness, rhythm, and timbre respectively correspond to amplitude, time, and both frequency and time properties. Pitch may be defined as "that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale" [18, 19]. (The pitch value to a sound is generally assigned by the frequency of a pure tone having the same subjective pitch as that sound [18].) While pitch or pitch value specifies the pitch sensation along a scale from low to high, other pitch parameters define additional pitch sensations independent of pitch value, e.g., pitch strength specifies the sensation along a scale from faint to distinct [20].

In the field of psychology of music, relations between musical features and humans' emotion and evaluation have been studied for decades. For example, high pitch, wide pitch range and large pitch variation may be associated with emotions like high activation, excitement, surprise [21], happiness [22, 23], pleasantness, anger, and fear [24]; low pitch and narrow pitch range may be associated with low activation, calmness, boredom, sadness [22], unpleasantness, and pleasantness, and small pitch variation with anger and fear [25].

(The apparent contradiction, e.g., both high and low pitch are associated with pleasantness, may depend on the context, that is, the combination and interaction with other features [25].) It is expected that these pitch parameters might be useful in soundscape measurement, especially for the emotional evaluation of soundscapes [26]. However, unlike music and speech, environmental sounds may be mainly composed of noise rather than discernable complex and/or pure tones, thus, there is a need to study the applicability of the pitch features to environmental sounds.

This paper, therefore, aims first to explore the pitch algorithms and parameters applicable to soundscape analysis, and then to study the pitch characteristics of various different environmental sounds. In the rest of this paper, first, the method for sound sample collection is described. Then, a number of existing algorithms are implemented with simplification/modification for environmental sounds. From these implemented models, the one with the best simulation performance for environmental sounds is selected, using a small size of sound samples. A number of pitch parameters, which correspond to subjective pitch sensations, and their statistical indices, which describe the variations of these parameters over time, are derived/developed based on the model selected. Finally, the characteristics of, and differences among various environmental sounds are studied in these pitch features, using a relatively large sample size.

2 Sound sample collection

Environmental sound sources that are often heard in outdoor soundscapes of everyday life are considered in this study, which include sounds from both nature and human activity/facility. The sound recordings were collected from multiple databases and supplemented by recordings made by the authors. Recordings were made in the countryside, natural parks, and urban areas in England, from 1994 to 2010. Calibration was based on measured SPLs, or in cases where calibration data was not available, based on reasonable estimates of the SPL range [14, 16]. The recordings are mono channel, 30 seconds in duration, and are sampled at 44,100 Hz (16 bit). Further details (including the recording equipment used) of the recordings can be found in [14].

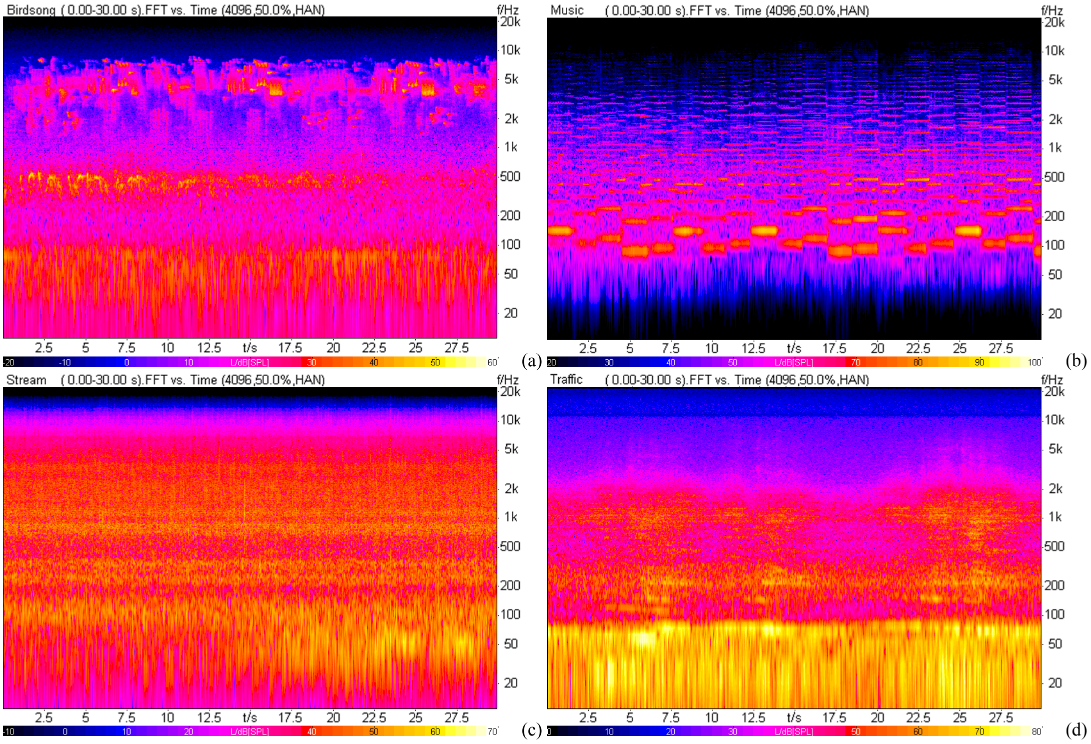


Fig. 1. Spectrogram: (a) birdsong, (b) music, (c) stream, and (d) traffic.

To examine the applicability of the pitch algorithms to soundscape research (Sections 3 and 4.1), a small set of samples are used, including 11 environmental sounds in which a single sound source is predominantly present and 4 soundscape sounds in which multiple sound sources are simultaneously present. The 11 environmental sound recordings are sounds of stream, river, sea waves, wind, birdsong, fountain, church bells, street music, street machinery, traffic, and voice. The 4 soundscape sound recordings are from 4 common urban places, and have different combinations of sound sources; they are soundscapes on a street with traffic, clock (Big Ben), and talking sounds, in a park with fountain and geese sounds, in a market place with talking sounds and footsteps, and in an urban square with music and talking sounds. Since the applicability of pitch algorithms is affected by the spectrum of audio signal to be estimated, specifically, whether it is composed of mainly broadband/band-pass noises or complex/pure tones, this small set of samples, from the collected recordings, is selected to cover a representative range of spectra and their variations of time of natural and urban environmental sounds [27]. For example, wind and traffic sounds are studied for broadband noises, machinery for band-pass noises/tones, birdsong, church bells, music, and voice for tones, and stream sounds for a combination of noises and tones. The natural water sounds of stream, river, and sea waves are all included since they have quite different patterns of variations of spectrum. The spectra over time of part of the recordings are shown in Fig. 1. It shows that the music in particular has a rather different spectrum compared to the other sounds, suggesting the importance of studying the applicability of musical features to environmental sounds.

To analyze the characteristics of various environmental sounds in terms of the pitch features (Section 5), 102 30-s recordings with single sound sources are used. Also, the correlation and principal component analyses of the developed pitch indices (Sections 4.2 and 4.3) are based on the large sample set. This set of recordings comprises natural sounds, which include water sounds (stream, river, and sea waves), wind sounds (in deciduous/coniferous trees and heathland), and birdsongs (in woodland, heathland/grassland, moorland/wetland, farmland, and coastal), and human activity/facility sounds in urban areas, which include sounds of church bells, fountains, street music, street machinery (e.g. cleaning machine, rubbish bin loading, and construction work), traffic, voice, and footsteps. In this paper, fountain sounds, differing from natural water sounds, are included in urban sounds, according to the definition of natural sounds in [16] (their primary excitation mechanism are not from nature), and since they have higher energy in high frequencies than natural water sounds. (More details about the composition of the 102 sound samples, including the numbers of recordings in each sound category, i.e., water sounds, wind sounds, birdsongs, and urban sounds, can be found in Fig. 4 and Table 5.)

3 Model implementation and comparison

A number of pitch models, i.e. spectral model and temporal model, according to the two classes of pitch perception theories, and simplification model for real-time pitch analysis in music and speech, are simplified/modified and implemented in a MATLAB program with MIRtoolbox 1.3.4, and then compared based on their pitch estimation performance for environmental sounds. The MIRToolbox is a MATLAB toolbox dedicated to the extraction of musically related features from audio signals, within the context of music information retrieval (MIR) [28].

3.1 Spectral model

There have been two classes of pitch perception theories that attempt to correlate the pitch of stimuli with the anatomical properties and physiological responses of the auditory system [18]. One of them is the 'place' theory [29]. As the spectral analysis taking part in the inner ear, different frequencies of a stimulus excite different places along the basilar membrane (BM) and hence neurones with different characteristic frequencies (CFs). The 'place' theory proposes that the pitch is determined by the recognition of excitation pattern of different places along the BM.

The implementation of the spectral model is based on Wightman’s mathematical "pattern-transformation model" [30], which shows a family similarity to Terhardt’s pattern recognition model [29, 31, 32] (a combination or competition of spectral-pitch pattern and virtual-pitch pattern) and Goldstein’s theory [33, 34], and is less computationally sophisticated. It is implemented through the calculation of cepstrum, which is defined as the power spectrum of the amplitude-logarithm of the power spectrum [35], and has been used for pitch detection in voiced-speech [36, 37]. That is, the model calculates first the power spectrum of an audio signal, which roughly transforms the stimulus into a pattern of peripheral neural activity or response of the BM, and then performs a Fourier transformation on the power spectrum. In other words, the output transformed pattern is the autocorrelation function of the signal; this is a spectrally based autocorrelation model (frequency domain computation) and thus phase-insensitive, which is different from temporally based autocorrelation models (time domain computation) as following in Section 3.2 that are phase-sensitive [38]. The pitches correspond to the peaks in the cepstrum: The abscissas of the function correspond to the reciprocal of the values of the perceived pitches, whereas the ordinates are related to the corresponding pitch strengths of the pitches [16].

3.2 Temporal model

An alternative to the place theory is the 'temporal' theory [18]. When a neurone is excited, the nerve firings tend to occur at a particular phase of the stimulating waveform, and thus the intervals between successive neural impulses approximate integral multiples of the period of the stimulating waveform (i.e., phase locking). The 'temporal' theory suggests the pitch is related to the temporal patterns of neural impulses within and across neurones.

The implementation is based on the temporal models of Moore [18] and Meddis et al. [39-41], with some simplifications for pitch simulation of environmental sounds in this paper. This simplified model consists of four key stages: (1) The signal is decomposed through a bank of critical-band filters that simulates the frequency analysis of the cochlea or BM [18, 38, 42]. The output of each filter corresponds to the mechanical motion of the BM and roughly represents the nerve impulses at that point. (2) Within each channel, an autocorrelation analysis is performed on the output filterband waveform. The autocorrelation estimates a distribution of time intervals among all spikes (or nerve fibre firing probabilities), similar to Licklider [38] and Meddis et al. [39-41]. This approximates the time intervals between only successive spikes [18] and is computationally convenient. (3) All the autocorrelation functions (ACFs) are averaged across channels to generate a summary autocorrelation function (SACF). (4) The peaks of the SACF are picked, which correspond to the pitches.

In stage (1), a number of different auditory filters are used, in order to compare their pitch simulation performances and find the one with the optimal performance for environmental sounds, in terms of computational accuracy and efficiency. These auditory filters include the gammatone filterbank [43], Bark scale critical bands [20], and third-octave band filters. They are all commonly used to represent the magnitude characteristic of the human auditory filter [43-45]. For gammatone filters, 10, 20, 40 and 80 filters (e.g. 60 or more gammatone filters have been used in Meddis et al.'s models [39-41]) with half overlapping along a scale between 50 and 22000 Hz are used here. The gammatone filterbank calls the Auditory Toolbox in MATLAB [46, 47]. Third-octave band filters have been used in loudness calculating procedure as an approximation of critical bands [20, 48]. The third-octave band filterbank used in this paper consists of 21 non-overlapping bands which cover the frequencies from 44 to 18000 Hz (the lowest three filters are one-octave band-pass filters) [49]. The temporal models using these different filterbanks are compared in Section 3.6.

3.3 *Simplification pitch model*

Based upon the pitch algorithms for simulation of auditory perception, some simplification pitch models have been developed for practical application of real-time pitch analysis in music and speech, and are thus computationally efficient and may be applicable for large sample size analysis. Part of Tolonen and Karjalainen [50]'s model is implemented, which can be seen as a computational simplification of the model of Meddis and O'Mard [41]. The procedure is as follows. (1) Instead of multi-channels in Section 3.2, this model essentially divides the signal into two channels, above and below 1000 Hz. (2) The envelope of the high-channel signal is then calculated by lowpass filtering at cut-off frequency of 1kHz. (3) The model computes ACFs of the low-channel signal and of the envelope of the high-channel signal. (4) The resulting two ACFs are averaged to produce a SACF, in which the peaks indicate the pitches of signal. The exponential magnitude compression of the "generalized" autocorrelation and the enhanced SACF in Tolonen and Karjalainen [50]'s model are not used in this study, since they increase the risk of sensitivity to noise and thus are not suitable for environmental sounds [50].

3.4 *Parameter setting*

The calculation parameters of these models implemented are set for environmental sounds based on the small set of sound samples. For each of these different models, unitary pitch range, between 75 and 5000 Hz, is considered. The lowest pitch value taken into consideration follows the convention in pitch estimation [51]. For the highest pitch value, above 4-5kHz the ability of the auditory system to discern changes in the frequency of pure tones diminishes and the sense of musical pitch disappears. Furthermore the tones produced by musical instruments, the human voice and most everyday sound sources all have fundamental frequencies below this range [18].

During peak selection from the cepstrum or SACF, the total amplitude of the function is firstly normalized between 0 and 1. It is postulated that a given local maximum of the function will be considered as a peak if it meets a number of conditions: (1) its normalized amplitude is higher than a threshold, specified by the parameter of "Threshold"; (2) the differences of amplitude with both the adjacent local minima are higher than a threshold, specified by the parameter of "Contrast"; (3) abscissa distance (frequency distance) to adjacent peak is greater than a given threshold, specified by the parameter of "Reso". The higher peak remains out of two adjacent peak candidates with a distance lower than the threshold. In addition, the peak near the place corresponding to the abscissa of zero of the function is ignored [30]. Of each model, different values of the condition parameters have been compared and set to obtain the optimal and balanced results based on the 15 sound samples, in that the major peaks of the function with positive absolute autocorrelation coefficients are generally selected [16]. While multiple pitches can be picked, the most prominent four pitches (if there are more than four), i.e. the ones with the highest pitch strengths corresponding to the highest four peaks, are extracted to reflect the different pitch properties of the sounds, given that some of the 15 sound samples show only zero to two pitches whereas some others show even eight to ten pitches using most of the models (as discussed in Section 3.5 below), and also, the strengths of additional pitches are small and thus less relevant. The parameter settings, indicated in the corresponding commands in MATLAB for each model, are shown in Table 1. It is noted that the authors have modified the program to a small extent to meet the needs in this study, and thus a number of the commands are not directly available in MIRtoolbox.

For the calculation of variation of pitches over time, the signal is first decomposed into successive frames of short duration, and pitches are calculated within each frame. The frame length of 46.4ms and hop length of 10ms are used according to Tolonen and Karjalainen [50].

Table 1 Pitch values and pitch strengths of four prominent average pitches of 11 environmental sounds with different models.

		Bird song	Church bells	Fountain	Machine	Music	River	Sea waves	Stream	Traffic	Voice	Wind
Cepstrum *p=mirpitch('folder','Cepstrum','Max',5000,'Threshold',0,'Contrast',0.2,'Reso','SemiTone','total',4)*	PV	3815	2610	4421	2329	557	3163	1002	1105	3574	882	4437
		1991	2158	3658	3907	2486	2316	1189	3990	2700	1463	3650
		1095	3990	2754	2575	1430	801	1078	2742	1770	4344	-
		-	512	3148	1825	491	605	2741	1427	1522	2094	-
	PA (1e+4)	8.781	2.777	6.875	2.672	2.769	8.574	7.095	1.363	2.906	1.615	2.734
		4.590	2.776	6.205	2.744	1.770	5.359	6.974	1.082	2.593	1.577	1.870
		2.310	2.380	5.547	1.487	1.707	4.567	6.942	0.722	1.711	1.509	-
		-	2.243	5.540	1.107	1.679	3.738	6.020	0.713	1.430	1.395	-
2Channels *p=mirpitch('folder','2Channels','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Generalized',2,'Enhanced',0,'total',4)*	PV	483	114	823	174	146	-	-	861	-	272	-
		85	101	266	94	-	-	-	119	-	-	-
		248	129	394	-	-	-	-	394	-	-	-
		102	939	-	-	-	-	-	-	-	-	-
	PA	1.689	2.096	0.847	0.551	1.275	-	-	0.689	-	0.230	-
		0.644	1.502	0.233	0.420	-	-	-	0.403	-	-	-
		0.571	1.454	0.225	-	-	-	-	0.280	-	-	-
		0.522	1.433	-	-	-	-	-	-	-	-	-
10 Gammatone *p=mirpitch('folder','Gammatone','Max',5000,'Threshold',0.4,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0,'total', 4)*	PV	3672	232	733	200	221	205	445	735	-	218	214
		1917	464	451	93	148	732	206	442	-	444	439
		457	101	1159	-	111	445	112	1161	-	111	-
		533	177	276	-	446	1189	720	274	-	154	-
	PA	1.914	1.942	1.008	1.185	1.634	0.758	1.149	1.407	-	1.403	1.209
		1.706	1.835	0.620	0.622	1.082	0.748	0.807	0.980	-	0.691	0.762
		1.316	1.769	0.601	-	1.044	0.747	0.336	0.763	-	0.581	-
		1.084	1.615	0.390	-	1.036	0.627	0.285	0.707	-	0.243	-
20 Gammatone *p=mirpitch('folder','Gammatone','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0,'total',4)*	PV	4017	205	739	104	147	105	153	923	-	152	-
		2011	102	360	-	111	153	206	745	-	103	-
		453	93	459	-	558	357	275	276	-	279	-
		491	473	274	-	209	275	355	365	-	-	-
	PA	1.581	1.830	0.258	0.226	0.955	0.253	0.443	0.554	-	0.695	-
		0.832	1.685	0.253	-	0.603	0.180	0.315	0.427	-	0.501	-
		0.796	1.576	0.245	-	0.269	0.111	0.261	0.351	-	0.226	-
		0.641	1.491	0.234	-	-0.097	0.072	0.254	0.322	-	-	-
40 Gammatone *p=mirpitch('folder','Gammatone','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0,'total',4)*	PV	4072	115	813	176	148	-	-	1064	-	276	-
		2027	102	204	93	-	-	-	402	-	-	-
		493	938	258	-	-	-	-	118	-	-	-
		556	84	134	-	-	-	-	286	-	-	-
	PA	8.618	7.529	0.634	0.770	4.897	-	-	1.550	-	0.857	-
		4.714	7.130	0.384	0.675	-	-	-	1.151	-	-	-
		4.616	7.027	0.302	-	-	-	-	0.509	-	-	-
		3.702	6.728	0.288	-	-	-	-	0.330	-	-	-
80 Gammatone *p=mirpitch('folder','Gammatone','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0,'total',4)*	PV	4050	115	812	177	148	-	-	1067	-	274	-
		2017	102	205	93	-	-	-	400	-	-	-
		491	939	268	-	-	-	-	119	-	-	-
		551	84	376	-	-	-	-	281	-	-	-
	PA	275	3652	37	17	3629	-	-	77	-	649	-
		158	3511	27	14	-	-	-	53	-	-	-
		146	3449	22	-	-	-	-	46	-	-	-
		122	3294	19	-	-	-	-	9	-	-	-
Bark *p=mirpitch('folder','Bark','Max',5000,'Threshold',0.3,'Contrast',0.15,'Reso','SemiTone','Compress',2,'Enhanced',0,'total',4)*	PV	4078	102	745	125	149	130	-	1067	1122	275	-
		490	115	381	-	-	-	-	117	248	-	-
		552	949	101	-	-	-	-	398	117	-	-
		2036	84	278	-	-	-	-	278	-	-	-
	PA	2.217	1.470	0.149	0.716	0.894	0.170	-	0.205	0.508	0.230	-
		1.309	1.457	0.085	-	-	-	-	0.190	0.354	-	-
		1.118	1.344	0.083	-	-	-	-	0.190	0.192	-	-
		1.077	1.283	0.077	-	-	-	-	0.069	-	-	-
Third-octave *p=mirpitch('folder','Klapuri','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0,'total',4)*	PV	4085	102	751	177	149	-	94	887	-	94	-
		2035	115	374	93	-	-	-	402	-	276	-
		490	84	472	-	-	-	-	118	-	-	-
		549	941	949	-	-	-	-	297	-	-	-
	PA	1.922	1.841	0.170	0.311	1.313	-	0.170	0.377	-	0.232	-
		1.072	1.735	0.161	0.293	-	-	-	0.256	-	0.175	-
		0.969	1.682	0.131	-	-	-	-	0.175	-	-	-
		0.887	1.607	0.107	-	-	-	-	0.093	-	-	-

Table 2 Pitch values and pitch strengths of four prominent average pitches of 4 soundscape sounds with different models.

		Street	Park	Market	Square
Cepstrum	PV	3142	2567	4017	3985
		2604	3682	3407	2970
		1288	1154	2311	-
		1001	2141	588	-
	PA (1e+4)	2.362	2.087	2.485	2.627
		1.591	1.766	1.498	2.576
		1.429	1.562	1.377	-
		1.246	1.443	0.839	-
2Channels	PV	99	77	-	87
		168	133	-	-
		-	91	-	-
		-	470	-	-
	PA	0.589	0.361	-	1.599
		0.426	0.358	-	-
		-	0.276	-	-
		-	0.269	-	-
10 Gammatone	PV	216	449	437	93
		99	721	223	206
		449	226	-	-
		-	114	-	-
	PA	1.532	1.366	1.444	0.984
		0.792	0.758	0.681	0.851
		0.147	0.679	-	-
		-	0.411	-	-
20 Gammatone	PV	102	153	-	98
		153	354	-	-
		-	272	-	-
		-	215	-	-
	PA	0.607	0.566	-	0.711
		0.341	0.479	-	-
		-	0.474	-	-
		-	0.352	-	-
40 Gammatone	PV	99	153	-	87
		169	77	-	-
		-	134	-	-
		-	91	-	-
	PA	2.268	1.516	-	7.127
		1.393	1.476	-	-
		-	1.410	-	-
		-	1.201	-	-
80 Gammatone	PV	99	77	-	87
		169	153	-	-
		552	134	-	-
		-	91	-	-
	PA	270	738	-	209
		195	697	-	-
		146	636	-	-
		-	618	-	-
Bark	PV	100	134	128	175
		169	152	-	93
		-	77	-	-
		-	90	-	-
	PA	0.674	0.359	0.244	0.638
		0.537	0.346	-	0.439
		-	0.333	-	-
		-	0.313	-	-
Third-octave	PV	99	151	-	87
		169	134	-	-
		-	469	-	-
		-	77	-	-
	PA	0.472	0.262	-	2.725
		0.276	0.260	-	-
		-	0.234	-	-
		-	0.192	-	-

3.5 Comparison of pitch models

To compare the simulation performance of the models implemented on environmental sounds, the pitch/pitches of the 15 sound samples are calculated according to each of the models. Both the variation of pitches over time and average pitches over the whole duration are calculated. The average pitch is calculated by ACF based on the whole duration, in contrast to ACFs of successive frames; the results of the sound samples using all the models, in terms of average pitch values (PV) and corresponding pitch strengths represented as amplitudes (PA), are shown in Tables 1 and 2. From the tables, it can be seen that the results are quite different across the different models, though some matches.

The simplification model (i.e. the '2Channels' method) has the limited frequency analysis range of pitch, focusing on low-to-mid fundamental frequencies, with maximum pitch of around 1kHz for the 15 sounds. It is determined by the boundary of the two channels, as both channels have the low-pass characteristics at the frequency of 1 kHz. Unlike most music and speech sounds, the pitches of environmental sounds may exceed that region as expected, e.g., the birdsongs may have pitches of about 4kHz using the temporal models as discussed in the following paragraphs. Since the two-channels method is a simplification of the temporal method, it is expected that the temporal models may derive more accurate results. Therefore, the two-channels pitch model may not be applicable for environmental sounds because of its limitation on pitch analysis range.

Using the spectral model based on the computation of cepstrum, the pitches of most of the 15 sounds are high. As shown in Tables 1 and 2, the most prominent pitches have values of above 2.5 kHz for the majority of the sounds (exceptions include pitch values of ~1 kHz for sea wave, stream, and voice, and ~500 Hz for music). A possible reason of these relatively high pitch results is that environmental sounds may consist of large amounts of noise, rather than pure or complex tones as in music. While Wightman [30]'s and cepstrum methods focus on the analysis of complex tones – the power spectra of which consist of evenly spaced components, for noises, the estimations of high pitch values may result from the random changes of noise signals and consequently the quick changes in spectra along the frequency scale, but may not correspond to real pitch sensation. Thus, the cepstrum method implemented can only be used for pitch analysis of certain sound types such as music and speech, but not for general environmental sounds. Although the inadequacy of the algorithm might be corrected with a number of modifications, e.g., a pre-whitening filter involved in Tolonen and Karjalainen [50]'s model to remove short-time correlation of the signal, no effort has been made to implement these modifications due to the complexity [30].

Using the temporal models, it can be seen from Tables 1 and 2 that the values of the pitches calculated vary among the 15 sounds. Taking the results by the algorithm based on 40 gammatone filter bands for example, the values of the most prominent pitches are ~4000 Hz for birdsongs, ~1000 Hz for sounds of fountain and stream, and no pitch perceived for river, sea waves, traffic, and wind. The pitch strengths of birdsongs, church bells (both above 7.5 for the most prominent pitches), and music (about 4.9) are higher than the others. These results may be consistent with what could be expected of human's pitch perception of environmental sounds; for example, traffic sounds consist of mainly broadband noise (without steep spectral slopes [20]) and thus do not evoke any pitch sensation. For sounds with harmonics, such as church bells and music, pitch values correspond to the fundamental frequencies. Indeed, the temporal theories/models were thought to explain two pitch perception mechanisms associated with both resolved and unresolved harmonics [18], and proved to be capable of explaining the majority of experimental results in pitch perception [42], including both complex tone and interrupted noise [39].

3.6 *Comparison of filterbanks*

The temporal models may be the appropriate methods for pitch analysis of environmental sounds as discussed above; however, it would be computationally expensive if a large number of filter bands were used. It is expected that the larger the number of filters used in this paper the more accurate the result would be, but computation time would also increase with increasing number of filters. Hence, the simulation performances based on the gammatone filters with different numbers of filter bands (10, 20, 40, and 80), the Bark scale filters, and the third-octave band filters are compared, in order to look for a balance between computational accuracy and efficiency.

As shown in Tables 1 and 2, between the gammatone filters, the pitch results differ when different numbers of filters are used. 40 and 80 gammatone filters generally produce similar results of pitch values in terms of variation of pitches over time, whereas the pitches over time computed with 10 or 20 gammatone filters tend to congregate more at certain frequencies. It is likely to be caused by the very limited number of filterbands. Both the Bark scale and third-octave band filters produce similar average pitch results to 40/80 gammatone filters, though the order of the four most prominent pitches – i.e. sorted by pitch strength from high to low – may vary. Also, these results (both pitches and relative pitch strengths) are somewhat similar to those produced by the '2Channels' method, except for high frequencies above about 1kHz. This agreement between the two types of models from another aspect supports the reliability of the pitch results.

In terms of calculation speed, both the Bark scale and third-octave band filters have similar numbers of filters to 20 gammatone filters and thus similar calculation speeds, all of which are quicker than the 40/80 gammatone filters. In other words, the Bark scale and third-octave band filters have similar accuracy performance to the gammatone filters with higher numbers of bands, but reduced computation time in this study. The calculation speed of the third-octave band filters is slightly quicker than the Bark scale filters because of its computational simplicity. Based on these results, therefore, the simplified temporal model implemented with the third-octave band filters is selected for the further pitch analysis of environmental sounds in this paper.

4 Determination of pitch parameters for environmental sounds

4.1 Pitch parameters based upon statistic analysis

Based on the model selected above, three pitch parameters related to subjective pitch sensations are derived. They are pitch value (PV), pitch strength (PA), and percentage of audible pitches over time (PN) (i.e. the ratio between the numbers of frames with pitch produced and the total frames in the duration). In order to describe the variations of pitch characteristics of sounds over time, a number of statistical indices are calculated from the results of the 15 sound samples.

The histograms of the four prominent pitches over time generally do not differ significantly for each sample, thus, to simplify the calculation, only the most prominent pitch (one pitch) in each frame is used. The histogram of the most prominent pitches over time of birdsongs is shown in Fig. 2, as an example. The histograms are non-normally distributed along the linear frequency scale. Therefore, to summarize the pitch data over time, a number of descriptive statistical indices are calculated for the pitch features (i.e. PV, PA and PN). For PV and PA, the statistical indices include average (AVE), median, mode (the value which occurs most frequently in the data), standard deviations (STDEV and STDEVA), maximum, minimum, range (the difference between the maximum and minimum), and 5%, 10%, 25%, 75%, 90%, and 95% percentiles.

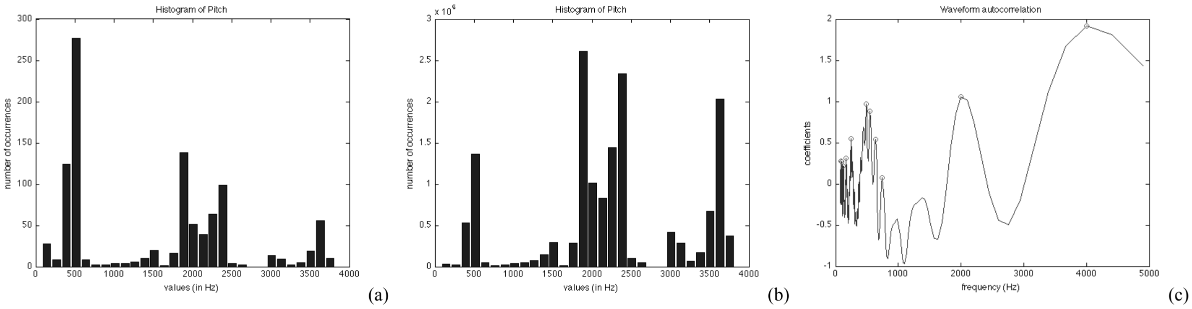


Fig. 2. Pitch statistics: (a) histogram, (b) weighted histograms, and (c) SACF.

Table 3 Component matrix and communalities of pitch indices.

	Component matrix				Communalities	
	PC 1	PC 2	PC 3	PC 4	Extraction of 3 PCs	Extraction of 4 PCs
PV1	**0.807**	0.165	0.200	-0.268	0.718	0.790
PV2	**0.628**	0.250	0.396	-0.158	0.614	0.639
PV3	**0.692**	0.259	-0.102	-0.129	0.556	0.573
PV4	**0.580**	0.074	0.061	0.373	**0.346**	**0.485**
PA1	**0.785**	-0.422	0.283	-0.024	0.874	0.875
PA2	**0.653**	**-0.664**	0.290	0.128	0.951	0.968
PA3	**0.684**	**-0.631**	0.299	0.117	0.956	0.970
PA4	**0.691**	**-0.589**	0.280	0.178	0.904	0.935
PN	**-0.826**	0.142	0.387	0.170	0.852	0.880
PV AVE	**0.939**	0.286	0.018	0.075	0.963	0.969
PV Mode	**0.665**	0.418	-0.296	0.446	0.705	0.904
PV STDEV	**0.835**	0.260	0.243	-0.301	0.825	0.915
PV STDEVA	**0.744**	0.437	0.398	0.050	0.904	0.906
PV Range	**0.591**	**0.513**	0.193	-0.339	0.650	0.764
PV Percentile5	**0.785**	0.281	-0.382	0.291	0.841	0.926
PV Percentile25	**0.835**	0.257	-0.100	0.294	0.773	0.859
PA AVE	**0.751**	-0.314	**-0.515**	-0.177	0.928	0.959
PA STDEV	**0.598**	-0.360	**-0.570**	-0.328	0.813	0.920
PA STDEVA	**0.802**	-0.208	-0.235	-0.046	0.741	0.743

As discussed in Section 3.5, another way for describing the average pitch of a given sound is calculating it from the ACF based on the whole duration, in contrast to averaging pitches over time. The results from the sound samples show the shapes of the SACFs are similar to those of the weighted histograms, see Fig. 2 as an example. Weighted histograms, which also take into account the strength of each pitch, are computed by adding the strengths of the pitches rather than counting the number of pitches in each frequency bin on a histogram.

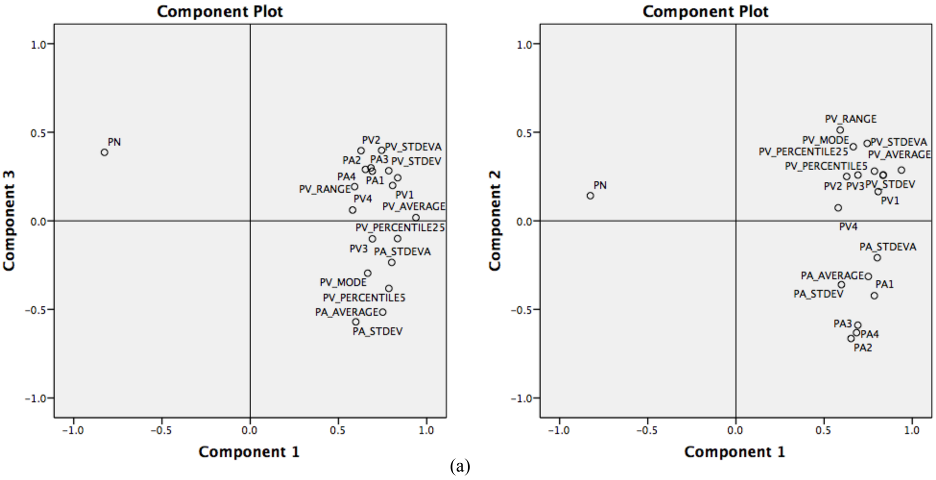
Thus, indices based on the SACF over the whole duration reflect also the characteristics of weighted histograms to some degree, and are used in this paper. These indices include values of the four prominent average pitches (PV1, PV2, PV3, PV4) and their pitch strengths or amplitudes (PA1, PA2, PA3, PA4). In total, 26 indices are included, part of which are shown in Table 3.

It is noted that more parameters and indices could be extracted based on the SACFs from which pitches are calculated or from the variation of pitches over time, e.g. pitch ambiguity, in addition to pitch strength indicated by the absolute height of peaks in the SACF pattern, which is thought to be related to the relative height and number of neighboring peaks in the pattern [30], and variance of successive pitches. However, this paper focuses on the indices discussed above.

4.2 Correlations and principal components of the pitch indices

Principal component analysis (PCA) is implemented to reduce the dimensionality of the dataset. The following analyses are based on the 102 samples, and the SPSS Statistics 20 software is used. Before the PCA, the correlations between the 26 pitch indices discussed above are first examined. The cases with missing values are excluded pairwise. The results shows that the correlations are high (coefficients above 0.8) between a number of the indices, e.g. between PA1, PA2, PA3, and PA4, between the statistical indices of PV over time, and between the statistical indices of PA over time. Among the indices, the ones that are particularly highly correlated (coefficients above 0.95) with some others are in general excluded from the index set for the PCA, for which 19 indices are remained.

The PCA is conducted on the correlation matrix of the 19 pitch indices. The cases with missing values are excluded listwise. Kaiser-Meyer-Olkin measure of sampling adequacy shows a result of 0.75, which generally indicates the adequacy of the sample size and the availability of the analysis. Among the 19 components extracted, the eigenvalues of first four components are greater than one. The first component accounts for 54.4% of the total variance, while the second, third and fourth account for 14.4%, 9.7% and 5.6% respectively. Table 3 shows the correlations between the first four components and indices, with high correlations (above 0.5) highlighted, and the proportion of each index's variance that can be explained by the retained principal components (PCs). It shows that all these indices (expect for PV4, of which the proportion is below 0.5) are generally well represented by the first four or three PCs. Component 1 represents almost all the indices as it has high correlations with them all, whereas Component 2 has high correlations with PA2, PA3, PA4, and PV Range, and Component 3 has high correlations with PA AVE and PA STDEV, but both mainly represent only a few of the indices. These results suggest that the pitch indices may form a single dimension of the variance based on the current dataset used in this study. These results can also be seen on the component loading plots, shown in Fig. 3, where the first three components are displayed. It can be seen that the indices are generally clustered in groups. In Fig. 3 (a), Component 1 mainly separates PN from the rest of the indices, whereas Component 2 separates PV and PA indices. In Fig. 3 (b), Component 3 approximately separates statistical indices of PV/PA over time and PV/PA of the four average pitches for the whole duration. (Exclusive PN, PCA generally generates the similar results for the PV and PA indices.)



(a) (b)

Fig. 3. Loading plot of the principal components of pitch indices, (a) Components 1 and 2, and (b) Components 1 and 3.

Table 4 Pearson's correlations between pitch and timbre indices.

	S AVE	S STDEV	S MAX	S MIN	Ton AVE	Ton STDEV	Ton MAX	Ton MIN	Fls AVE	Fls STDEV	Fls MAX	Fls MIN
PV1	0.395**	**0.620****	**0.654****	-0.017	-0.035	-0.003	0.168	-0.086	0.419**	0.483**	0.508**	0.205
PV2	0.411**	0.520**	**0.638****	0.056	-0.043	-0.076	0.114	c	0.392**	0.441**	0.489**	0.244
PV3	0.327*	0.480**	0.491**	-0.068	-0.074	-0.109	0.079	c	0.200	0.294*	0.330*	-0.029
PV4	0.319*	0.290*	0.352*	0.044	0.021	0.025	0.102	c	0.321*	0.166	0.273	0.291*
PA1	0.156	**0.643****	0.539**	-0.262*	0.382**	0.480**	0.587**	0.090	0.517**	0.533**	0.559**	0.339**
PA2	0.003	0.503**	0.418**	-0.348**	0.552**	0.580**	**0.646****	c	0.426**	0.439**	0.474**	0.282*
PA3	0.039	0.492**	0.400**	-0.322*	0.563**	0.580**	**0.627****	c	0.367**	0.389**	0.415**	0.178
PA4	0.064	0.470**	0.385**	-0.285	0.546**	0.555**	0.570**	c	0.354*	0.362*	0.378**	0.163
PN	-0.192	-0.433**	-0.422**	0.149	-0.134	-0.241*	-0.431**	0.072	-0.408**	-0.393**	-0.424**	-0.268**
PV AVE	0.578**	**0.759****	**0.792****	0.081	0.051	0.100	0.292**	-0.070	**0.623****	**0.602****	**0.641****	0.431**
PV Mode	0.591**	**0.611****	**0.683****	0.166	0.051	0.079	0.252*	-0.044	0.571**	0.446**	0.516**	0.507**
PV STDEV	0.545**	**0.771****	**0.778****	0.059	0.028	0.101	0.290**	-0.098	**0.620****	**0.624****	**0.661****	0.444**
PV STDEVA	0.565**	**0.740****	**0.769****	0.111	0.017	0.060	0.213*	-0.087	**0.615****	**0.642****	**0.677****	0.405**
PV Range	**0.612****	**0.646****	**0.715****	0.236*	-0.062	0.018	0.116	-0.146	0.533**	0.488**	0.518**	0.429**
PV Percentile5	0.568**	**0.701****	**0.696****	0.076	0.074	0.129	0.291**	-0.046	0.597**	0.471**	0.513**	0.496**
PV Percentile25	0.516**	**0.691****	**0.714****	0.070	0.060	0.097	0.249*	-0.051	0.555**	0.524**	0.552**	0.372**
PA AVE	0.419**	**0.690****	**0.652****	-0.086	0.338**	0.449**	**0.633****	0.096	**0.637****	0.520**	0.573**	0.546**
PA STDEV	0.342**	0.575**	0.534**	-0.089	0.274**	0.402**	0.550**	0.034	0.507**	0.403**	0.446**	0.437**
PA STDEVA	0.405**	**0.739****	**0.716****	-0.118	0.426**	0.547**	**0.708****	0.095	**0.769****	**0.705****	**0.766****	0.598**

** and * respectively indicate correlation is significant at the 0.01 level and 0.05 level (2-tailed), and c indicates it cannot be computed because at least one of the variables is constant.

4.3 Correlations between the pitch indices and the timbre indices

To check if any of the pitch indices represents the same or similar variance with the other psychoacoustic indices, including loudness and timbre (sharpness, tonality, roughness, and fluctuation strength), that have been used for soundscape analysis [14], the correlations between the 19 pitch indices and these previous psychoacoustic indices are examined. The cases with missing values are excluded pairwise. The results show that the correlation coefficients generally are not very high (below 0.6), with the highest coefficients of 0.6 to 0.8 existing between certain pitch and timbre indices; e.g. between pitch (both value and strength) and variation and maximum of sharpness (S STDEV, S MAX), between pitch strength and maximum tonality (Ton MAX), and between pitch (both value and strength) and fluctuation strength (Fls AVE, Fls STDEV, Fls MAX); parts of the results are shown in Table 4. In other words, the pitch indices developed in this paper in general provide additional variance to the psychoacoustic indices that have been used in soundscape analyses. These correlations between pitch and timbre indices can be understood in the way that either there are certain inherent common variances contained in the parameters or indices, or the correlations appear based on the current data set, i.e., certain samples show a number of characteristics simultaneously, e.g., the birdsong recordings have high pitch value and strength and meanwhile high sharpness, tonality and fluctuation strength.

5 Pitch characteristics of environmental sounds

5.1 Hierarchical cluster analysis

With the 19 pitch indices discussed above, the 102 recordings are clustered gradually using hierarchical cluster analysis (HCA) that starts with each case in a separate cluster and then combines clusters until only one is left. The dendrogram is shown in Fig. 4. It shows that between the last two clusters, one has most of the birdsongs recordings. In the other, there are a sub-cluster of church bells and a sub-cluster of some music, voice and birdsongs; the categories are rather mixed in the other sub-clusters. The different characteristics of the sounds in different categories in terms of the pitch indices are analyzed in the following section with one-way analysis of variance (ANOVA).

5.2 One-way analysis of variance

The mean values of the four sound categories, i.e. water, wind, birdsongs, and urban sounds, in terms of the 19 pitch indices, are compared with ANOVA, in order to examine if the sound categories differ from each other significantly in one or more indices. For each index, the cases with missing values are excluded from the analysis. The assumption of ANOVA, i.e., the homogeneity of variances of the indices, is firstly tested. It shows that the p values of all the indices are less than the level of 0.05, which suggests the assumption is rejected and the variances are unequal.

Table 5 Means of pitch indices for the four categories.

		Water		Wind		Bird		Urban	
	Total	N	Mean	N	Mean	N	Mean	N	Mean
PV1	75	24	232	9	530	28	2593	14	289
PV2	62	19	276	3	261	28	1030	12	249
PV3	50	14	199	0	-	27	1055	9	178
PV4	47	12	294	0	-	26	860	9	304
PA1	75	24	0.238	9	0.242	28	1.434	14	0.800
PA2	62	19	0.172	3	0.132	28	0.872	12	0.599
PA3	50	14	0.142	0	-	27	0.698	9	0.528
PA4	47	12	0.109	0	-	26	0.616	9	0.443
PN	102	34	0.969	23	0.816	28	0.352	17	0.649
PV AVE	102	34	233	23	202	28	1754	17	221
PV Mode	102	34	103	23	102	28	1313	17	128
PV STDEV	102	34	204	23	152	28	872	17	176
PV STDEVA	102	34	206	23	155	28	880	17	153
PV Range	102	34	2085	23	1163	28	3534	17	1513
PV Percentile5	102	34	83	23	83	28	466	17	87
PV Percentile25	102	34	105	23	101	28	1092	17	114
PA AVE	102	34	0.270	23	0.279	28	1.050	17	0.513
PA STDEV	102	34	0.081	23	0.102	28	0.557	17	0.241
PA STDEVA	102	34	0.088	23	0.110	28	0.447	17	0.240

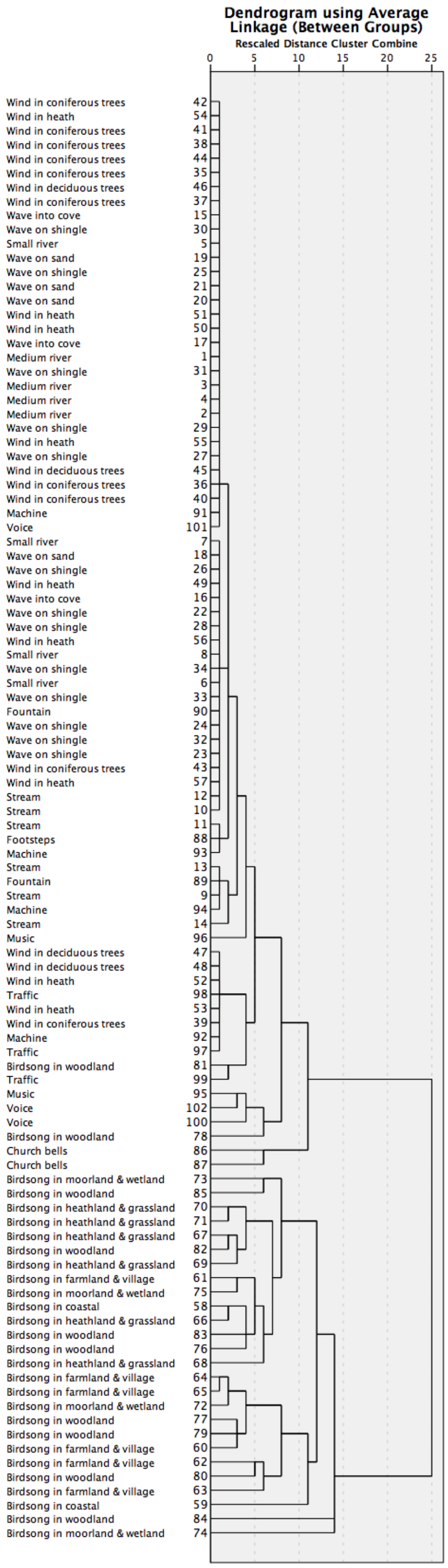


Fig. 4. Dendrogram for the 102 recordings by HCA.

The results of ANOVA show that the significance of F ratio (p value) is less than the level of 0.05 in terms of all the indices, suggesting some significant differences among the means of the categories, or between at least two categories. Furthermore, considering the inequality of the variances, post hoc tests are employed both to verify the results and to identify which of the specific categories differ. Table 5 shows the number of cases and the group means of each category. Table 6 shows the results of the post hoc tests using the Dunnett's T3 method, displaying the multiple comparisons among the categories in terms of difference between the means. Here, post hoc tests are not performed for PV3, PV4, PA3, and PA4, since the wind sound group has no cases showing result in the indices. The analysis shows that for almost all these pitch indices there are significant mean differences between birdsongs and the other three sound categories. Birdsongs have higher pitch values and pitch strengths, and lower percentage of audible pitches over time than the other three categories. In addition, a number of indices also show differences between these three other categories, as shown in Table 6. For example, between water and wind sounds, water sounds have a lower mean of PV1 and a higher mean of PV Range than wind sounds. More detailed results about the characteristics of the environmental sounds are discussed in the following section (Section 5.3).

5.3 Characteristics of the sound categories

Since high correlations exist among each group of indices as discussed above in Section 4.2, a number of key pitch indices among the 19 pitch indices used in the ANOVA can be are selected - PV1, PA1, PN, PV AVE, and PA AVE. They respectively represent the different index groups: Pitch values of the four average pitches for the whole duration, pitch strengths of the four average pitches, the percentage of audible pitches over time, statistical indices of pitch values over time, and indices of pitch strengths over time. They in general contribute most to the first component by PCA as shown in Table 3. With the 5 key indices, the 102 sound samples are plotted in the two-dimensional coordinate systems with their axes presenting the key indices, as shown in Fig. 5.

From Fig. 5, as well as Table 5 in Section 5.2, it can be seen that, as per the results of ANOVA, the recordings in birdsongs category have relatively high pitch values and pitch strengths compared to the other three categories. The values are generally above 1000Hz for PV1 and 500Hz for PV AVE, and above 0.5 for PA1 and PA AVE. Recordings in water and wind sound categories are located in almost the same areas in the figures; both have relatively low pitch values and strengths, generally below 1000Hz for PV1 and 500Hz for PV AVE, and below 0.5 for PA1 and PA AVE. In water sounds, PV AVE values of stream sounds are between 280 and 510Hz, whereas of river and sea waves sounds are below 280Hz. Recordings in the urban sound category generally have relatively low pitch values (below 1000Hz for PV1 and 500Hz for PV AVE) and a relatively wide range of pitch strengths compared to water and wind sounds, varying between about 0 to 2 for PA1 and about 0 to 1 for PA AVE (about 0.9 to 1.1 for music, 0.5 to 0.9 for voice, and generally below 0.5 for the other sounds). In terms of PN, it shows that the PN values of all birdsongs are relatively low, generally below 0.8; those of all water sounds are high, close to 1; and those of wind sounds and a majority of urban sounds are in a wide range, from about 0 to 1. Among urban sounds, the PN values of traffic sounds are low, about 0.08 to 0.15, and of church bells and fountain sounds are high, about 1. In other words, in general, birdsongs have fewer pitches audible over time than wind sounds, and than water sounds, which have audible pitches nearly throughout their duration. This could be explained by the nearly silent periods between successive birdsongs, whereas wind and water sounds are relatively constant. It is also noted that the PN index reflects a relative value in this paper. It is the percentage of relative audible pitches; when a sound sample, such as birdsongs, has pitches with very high pitch strength, other pitches in it (such as pitches in the nearly silent periods) become less notable and are not counted in the percentage.

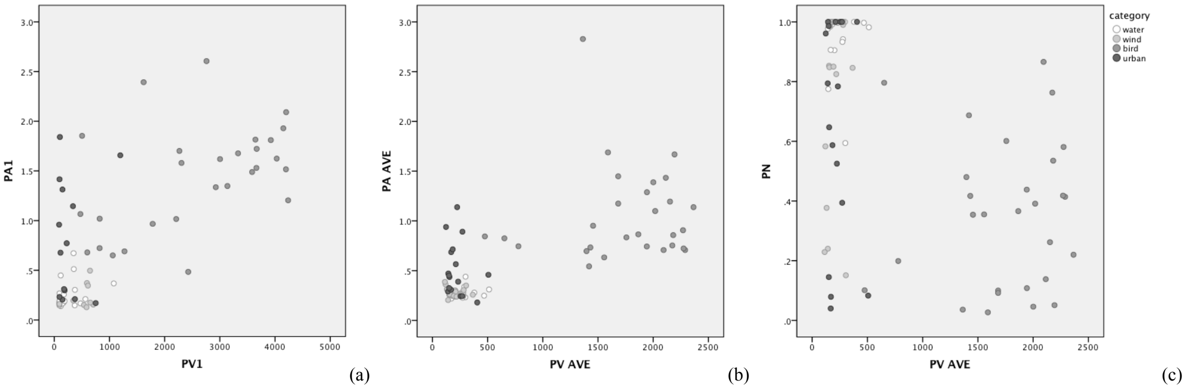


Fig. 5. Characteristics of the four sound categories in terms of the key pitch indices, (a) PV1 and PA1, (b) PV AVE and PA AVE, and (c) PV AVE and PN.

Table 6 Multiple comparisons (I-J) of pitch indices for the four categories.

(I) Category	Water			Wind			Bird			Urban		
(J) Category	Wind	Bird	Urban	Water	Bird	Urban	Water	Wind	Urban	Water	Wind	Bird
PV1	-298*	-2361*	-56	298*	-2063*	242	2361*	2063*	2305*	56	-242	-2305*
PV2	15	-754*	27	-15	-769	13	754*	769	781*	-27	-13	-781*
PV3	-	-	-	-	-	-	-	-	-	-	-	-
PV4	-	-	-	-	-	-	-	-	-	-	-	-
PA1	-0.004	-1.196*	-0.562*	0.004	-1.192*	-0.558*	1.196*	1.192*	0.633*	0.562*	0.558*	-0.633*
PA2	0.040	-0.700*	-0.427	-0.040	-0.739*	-0.467*	0.700*	0.739*	0.272	0.427	0.467*	-0.272
PA3	-	-	-	-	-	-	-	-	-	-	-	-
PA4	-	-	-	-	-	-	-	-	-	-	-	-
PN	0.153	0.618*	0.321*	-0.153	0.465*	0.168	-0.618*	-0.465*	-0.297*	-0.321*	-0.168	0.297*
PV AVE	31	-1521*	12	-31	-1552*	-19	1521*	1552*	1533*	-12	19	-1533*
PV Mode	1	-1210*	-24	-1	-1211*	-25	1210*	1211*	1185*	24	25	-1185*
PV STDEV	52	-667*	29	-52	-720*	-24	667*	720*	696*	-29	24	-696*
PV STDEVA	51	-674*	53	-51	-724*	2	674*	724*	726*	-53	-2	-726*
PV Range	922*	-1449*	572	-922*	-2371*	-350	1449*	2371*	2021*	-572	350	-2021*
PV Percentile5	0	-383*	-4*	0	-382*	-4	383*	382*	379*	4*	4	-379*
PV Percentile25	4	-987*	-9	-4	-991*	-13	987*	991*	978*	9	13	-978*
PA AVE	-0.009	-0.780*	-0.243*	0.009	-0.771*	-.0234*	0.780*	0.771*	0.537*	0.243*	0.234*	-0.537*
PA STDEV	-0.021	-0.476*	-0.159*	0.021	-0.456*	-.0139*	0.476*	0.456*	0.317*	0.159*	0.139*	-0.317*
PA STDEVA	-0.021	-0.359*	-0.151*	0.021	-0.337*	-0.130	0.359*	0.337*	0.207*	0.151*	0.130	-0.207*

* indicates significantly different group means at the 0.05 level.

6 Conclusions and discussions

By examining the pitch simulation performance of the different pitch models implemented in this paper, including temporal models and spectral models based on the pitch perception and simplified model in music and speech, the temporal model implemented is found to be applicable to the pitch analysis of the common environmental sounds in soundscapes considered. This simplified temporal model, based on the temporal theories/models of pitch perception, is implemented by the decomposition through third-octave band filters, autocorrelation computation, and pitch selection.

Using this model, a number of parameters that correspond to pitch sensations, i.e. pitch value, pitch strength, and the percentage of audible pitches over time, and statistical indices that describe the pitch features over time are developed for the analysis of environmental sounds in this study. The correlations between the pitch indices and the loudness and timbre indices that have been used in soundscape analyses are not very high, except for certain correlations (coefficients of about 0.6 to 0.8), e.g. between pitch (both value and strength) and sharpness, and between pitch strength and tonality. It suggests that the pitch indices developed provide additional variance to the previous psychoacoustic indices for soundscape measurement.

In terms of these pitch indices, the different characteristics of different environmental sounds are shown. In general, both water sounds and wind sounds have low pitch values and pitch strengths, and high percentage of audible pitches over time. Birdsongs have high pitch values and pitch strengths, higher than the other three categories, and low percentages of audible pitches. Urban sounds have low pitch values and a wide range of pitch strengths; they have higher mean pitch strength and lower mean percentage of audible pitches than water and wind sounds. Among urban sounds, pitch strengths are high for music and voice, and low for the other sounds. The percentages of audible pitches of traffic sounds are low, as they in general do not evoke any pitch sensation, and those of church bells and fountain sounds are high.

To certain degree, the results correspond to those on the association of pitch features and emotions in music (as discussed in Section 1). For example, birdsongs (high pitch) are usually perceived/evaluated as activation and pleasantness, whereas water and wind sounds (low pitch) are usually perceived as calmness and pleasantness (such as in [9]). Relationships between the pitch sensations as indicated in this paper and the emotional evaluations of soundscapes could be examined in future studies. Moreover, according the results in this paper on different pitch characteristics of various environmental sounds, the pitch parameters/indices could be used for the automatic recognition of environmental sound sources [15], such as to identify birdsongs from other sound sources. While this paper focuses mainly on the environmental sounds in which a single sound source is predominantly present, it is expected that the pitch simulation model is also applicable to more general soundscape sounds, since this algorithm is found to be suitable for the environmental sounds covering a wide range of spectra, with both noise and tone components.

Acknowledgments

This work was supported by the UK-China Scholarships for Excellence Scheme, jointly funded by the China Scholarships Council (CSC), the UK Department for Business, Innovation and Skills (BIS), and the University of Sheffield. The authors would like to thank the Positive Soundscape Project for providing the urban sound recordings, and Joe Kang for proofreading the manuscript.

References

[1] R.M. Schafer, The Tuning of the World, Knopf, New York, 1977.
[2] J. Kang, Numerical modelling of the sound fields in urban streets with diffusely reflecting boundaries, Journal of Sound and Vibration, 258 (2002) 793-813.
[3] E. Ohrstrom, A. Skanberg, H. Svensson, A. Gidlof-Gunnarsson, Effects of road traffic noise and the benefit of access to quietness, Journal of Sound and Vibration, 295 (2006) 40-59.
[4] W. Yang, J. Kang, Acoustic comfort evaluation in urban open public spaces, Applied Acoustics, 66 (2005) 211-229.
[5] T. Stockfelt, Sound as an existential necessity, Journal of Sound and Vibration, 151 (1991) 367-370.
[6] B. De Coensel, D. Botteldooren, T.D. Muer, B. Berglund, M.E. Nilsson, P. Lercher, A model for the perception of environmental sound based on notice-events, The Journal of the Acoustical Society of America, 126 (2009) 656-665.
[7] W.J. Davies, M.D. Adams, N.S. Bruce, R. Cain, A. Carlyle, P. Cusack, D.A. Hall, K.I. Hume, A. Irwin, P. Jennings, M. Marselle, C.J. Plack, J. Poxon, Perception of soundscapes: An interdisciplinary approach, Applied Acoustics, 74 (2013) 224-231.
[8] M. Raimbault, C. Lavandier, M. BÈrengier, Ambient sound assessment of urban environments: field studies in two French cities, Applied Acoustics, 64 (2003) 1241-1256.
[9] W. Yang, J. Kang, Soundscape and sound preferences in urban squares: A Case Study in Sheffield, Journal of Urban Design, 10 (2005) 61-80.
[10] D. Botteldooren, B. De Coensel, T. De Muer, The temporal structure of urban soundscapes, Journal of Sound and Vibration, 292 (2006) 105-123.

[11] M. Yang, B.D. Coensel, J. Kang, Presence of 1/f noise in the temporal structure of psychoacoustic parameters of natural and urban sounds, The Journal of the Acoustical Society of America, 138 (2015) 916-927.

[12] K. Genuit, A. Fiebig, Psychoacoustics and its benefit for the soundscape approach, Acta Acustica united with Acustica, 92 (2006) 952-958.

[13] A. Fiebig, V. Acloque, S. Basturk, M. Di Gabriele, M. Horvat, M. Masullo, R. Pieren, K.S. Voigt, M. Yang, K. Genuit, B. Schulte-Fortkamp, Education in soundscape - A seminar with young scientists in the COST Short Term Scientific Mission "Soundscape - Measurement, Analysis, Evaluation", Proceedings of the 20th International Congress on Acoustics (ICA), Sydney, Australia, 2010.

[14] M. Yang, J. Kang, Psychoacoustical evaluation of natural and urban sounds in soundscapes, The Journal of the Acoustical Society of America, 134 (2013) 840-851.

[15] M. Yang, J. Kang, Applicability and application of music features in soundscape, Proceedings of AIA-DAGA, Merano, Italy, 2013.

[16] M. Yang, Natural and Urban Sounds in Soundscapes, Ph.D. Thesis, School of Architecture, The University of Sheffield, Sheffield, 2013.

[17] M. Yang, J. Kang, Soundscape analysis using musical features with music information retrieval techniques, Proceedings of European Acoustics Association (EAA) 6th Forum Acusticum, Aalborg, Denmark, 2011, pp. 2025-2030.

[18] B.C.J. Moore, An Introduction to the Psychology of Hearing, 4th ed., Academic Press, London, 1997.

[19] American Standards Association, ASA Acoustical Terminology SI, 1–1960, New York, 1960.

[20] E. Zwicker, H. Fastl, Psychoacoustics – Facts and Models, Springer, Berlin, 1999.

[21] K.B. Watson, The nature and measurement of musical meanings, Psychological Monographs, 54 (1942) 1-43.

[22] K. Hevner, The affective value of pitch and tempo in music, American Journal of Psychology, 49 (1937) 621-630.

[23] L.L. Balkwill, W.F. Thompson, A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues, Music Perception, 17 (1999) 43-64.

[24] C.L. Krumhansl, An exploratory study of musical emotions and psychophysiology, Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Experimentale, 51 (1997) 336-353.

[25] A. Gabrielsson, E. Lindström, The influence of musical structure on emotional expression, in: P.N. Juslin, J.A. Sloboda (Eds.) Music and Emotion Theory and Research, Oxford University Press, New York, 2001, pp. 223-248.

[26] B. Schulte-Fortkamp, B.M. Brooks, W.R. Bray, Soundscape: An approach to rely on human perception and expertise in the post-modern community noise era, Acoustics Today, 3 (2007) 7-15.

[27] A.L. Brown, J. Kang, T. Gjestland, Towards standardization in soundscape preference assessment, Applied Acoustics, 72 (2011) 387-392.

[28] O. Lartillot, P. Toiviainen, A Matlab toolbox for musical feature extraction from audio, Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, 2007, pp. 237-244.

[29] E. Terhardt, Pitch, consonance, and harmony, The Journal of the Acoustical Society of America, 55 (1974) 1061-1069.

[30] F.L. Wightman, The pattern-transformation model of pitch, The Journal of the Acoustical Society of America, 54 (1973) 407-416.

[31] E. Terhardt, Calculating virtual pitch, Hearing Research, 1 (1979) 155-182.

[32] E. Terhardt, G. Stoll, M. Seewann, Algorithm for extraction of pitch and pitch salience from complex tonal signals, Journal of the Acoustical Society of America, 71 (1982) 671-678.

[33] J.L. Goldstein, An optimum processor theory for the central formation of the pitch of complex tones, Journal of the Acoustical Society of America, 54 (1973) 1496-1516.

[34] E. de Boer, Pitch theories unified, in: E.F. Evans, J.P. Wilson (Eds.) Psychophysics and Physiology of Hearing, Academic, London, 1977, pp. 323-334.

[35] B.P. Bogert, M.J.R. Healy, J.W. Tukey, The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking, M. Rosenblatt (Ed.) Proceedings of the Symposium on Time Series Analysis, John Wiley & Sons Inc., New York, 1963, pp. 209-243.

[36] A.M. Noll, Short-time spectrum and "cepstrum" techniques for vocal-pitch detection, The Journal of the Acoustical Society of America, 36 (1964) 296-302.

[37] A.M. Noll, Cepstrum pitch determination, The Journal of the Acoustical Society of America, 41 (1967) 293-309.

[38] J.C.R. Licklider, A duplex theory of pitch perception, Experientia, 7 (1951) 128-133.

[39] R. Meddis, M.J. Hewitt, Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification, The Journal of the Acoustical Society of America, 89 (1991) 2866-2882.

[40] R. Meddis, M.J. Hewitt, Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity, The Journal of the Acoustical Society of America, 89 (1991) 2883-2894.

[41] R. Meddis, L. O'Mard, A unitary model of pitch perception, The Journal of the Acoustical Society of America, 102 (1997) 1811-1820.

[42] B.C.J. Moore, Effects of relative phase of the components on the pitch of three-component complex tones, in: E.F. Evans, J.P. Wilson (Eds.) Psychophysics and Physiology of Hearing, Academic Press, London 1977, pp. 349-358.

[43] R.D. Patterson, I. Nimmo-Smith, J. Holdsworth, P. Rice, Spiral vos Final Report, Part A: The Auditory Filterbank, Contract Report (APU 2341), Cambridge Electronic Design, 1988.

[44] R.D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, M. Allerhand, Complex sounds and auditory images, in: Y. Cazals, L. Demany, K. Horner (Eds.) Auditory Physiology and Perception, Proceedings of the 9th International Symposium on Hearing, Pergamon, Oxford, 1992, pp. 429-446.

[45] R.D. Patterson, B.C.J. Moore, Auditory filters and excitation patterns as representations of frequency resolution, in: B.C.J. Moore (Ed.) Frequency Selectivity in Hearing, Academic, London, 1986, pp. 123-177.

[46] M. Slaney, An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank, Apple Computer Technical Report, 1993.

[47] O. Lartillot, MIRtoolbox 1.3.4 User's Manual, 2011.

[48] International Organization for Standardization, ISO 532:1975, Acoustics—Method for Calculating Loudness Level International Organization for Standardization, Geneva, Switzerland, 1975.

[49] A. Klapuri, Sound onset detection by applying psychoacoustic knowledge, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 06, Phoenix, Ariz., US, 1999, pp. 3089-3092.

[50] T. Tolonen, M. Karjalainen, A computationally efficient multipitch analysis model, IEEE Transactions on speech and audio processing, 8 (2000) 708-716.

[51] P. Boersma, D. Weenink, Praat: Doing Phonetics by Computer, http://www.fon.hum.uva.nl/praat/ (accessed 11.03.2016).