The University Of Sheffield.

Natural and Urban Sounds in Soundscapes

Ming Yang

July 2013

A thesis submitted to the University of Sheffield in partial fulfilment of the requirements for the degree of Doctor of Philosophy

School of Architecture
University of Sheffield

Abstract

Among various sounds in the environment, natural sounds, such as water sounds and birdsongs, have proven to be highly preferred by humans, but the reasons for these preferences are not yet completely understood. This research study explores the differences between various natural and urban environmental sounds from the viewpoint of objective measures. Moreover, since numerous studies of soundscape perception and evaluation have revealed that besides the conventional parameters, e.g., A-weighted sound pressure level, additional parameters are necessary for soundscape measurement, in this study more possible parameters are explored. From alternative algorithms of the features proposed in literature for both perception of the auditory system and practical application in music and speech, the algorithms applicable for environmental sound are searched through comparison.

The sound samples used in this study include the recordings of single sound source categories of water, wind, birdsongs, and urban sounds including street music, mechanical sounds and traffic noise. The samples are analysed with a number of objective parameters in three aspects, which include psychoacoustic parameters that have been recommended in previous soundscape researches, additional psychoacoustically related parameters that have previously mainly been applied in music perception, and 1/f noise dynamic that has been observed in music, speech, and soundscapes.

Based on one-way analysis of variance, hierarchical cluster, and principal components analyses of the calculated results, a series of differences are shown among different sound types in terms of key parameters, which include fluctuation strength, pitch, loudness, and 1/f noise. Generally, both water and wind sounds have low fluctuation strength, pitch values, and pitch strengths; birdsongs have high fluctuation strength, pitch values, and pitch strength, low loudness, and exhibit generally 1/f behaviour of loudness in short and medium time intervals; and urban sounds have low pitch values, high loudness, and relatively wide ranges of other parameters.

With the parameters, furthermore, the sound categories of recordings are automatically identified/classified using discriminant function analysis and artificial neural networks. With the artificial neural networks, which have better performance than the discriminant functions for the identification, based on all the psychoacoustic, music, and 1/f noise indices, the prediction accuracies are above about 99% for the three natural sound categories, i.e., of water, wind, birdsongs, and about 90% for the urban sound category.

Acknowledgements

I would like to express my deepest appreciation to my supervisor, Professor Jian Kang, for his indispensable guidance and support, with his specialist knowledge, phenomenal intellect and limitless patience.

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. Also, it was supported by the Short-Term Scientific Missions (STSM) through COST Action TD0804 on Soundscape of European Cities and Landscapes, and by STSM through COST Action IC0601 on Sonic Interaction Design (SID).

I would also like to thank the Positive Soundscape Project for providing the urban sound recordings. I would like to greatly appreciate Dr. Bert De Coensel for the computer program of 1/f noise calculation and useful discussions. I also appreciate the editor and reviewers of the publication of some part of this work for their valuable comments/suggestions.

I would like to extend my thanks to the acoustics group, the staff, and my friends in the School of Architecture, the University of Sheffield. I am grateful for their encouragement and support, and I enjoyed the time having spent with them.

I would like to specially thank Mr. Junjie Huang and my family for their understanding and support, which I could never have completed this work without.

Contents

List of tables

List of figures

Chapter 1
Introduction

1.1 Research Background

In traditional environmental noise control, considerable efforts have been made to reduce sound level, but it is gradually noticed that this does not necessarily lead to a better acoustic comfort (Kang 2006). Soundscape, a more positive way to deal with the environment noise pollution problem, has got much attention in the last few years. It was initially proposed by Schafer, a musician, in the 1960s, concerning not only the sonic environment but also the perception of humans (Schafer 1977; Truax 1999). Like the Bauhaus in Germany in the 1920s that invented the subject of industrial design and brought aesthetics to machinery and mass production, soundscape study intended to found an interdiscipline named acoustic design, in which the world is treated as a macrocosmic musical composition and to be designed (Schafer 1977).

Soundscape attempts to draw the independent areas of sonic studies together to answer the shared question that "the relationship between man and the sounds of his environment" (Schafer 1977). As different from traditional noise control, soundscape concerns not only sound, but regards sound, environment and people as a whole. This nature has brought together researchers from a wide range of academic backgrounds and created exchanges between the different disciplines (Schafer 1977; Karlsson 2000; Kang 2006). Using physical, linguistic, psychological and sociological etc. approaches to the soundscape studies respective or combined, soundscape evaluations have been intensively studied (Raimbault and Dubois 2005; Kang 2006), as well as the corresponding research methods such as survey, categorization, and auralisation (Dubois *et al.* 2006; Kang 2006). These researches showed that soundscape evaluations are influenced by physical, physiological, psychological, sociological and other factors along all aspects of sound, environment and people (Botteldooren and Verkeyn 2002; Raimbault and Dubois 2005; Genuit and Fiebig 2006; Kang 2006; Schulte-Fortkamp and Fiebig 2006; Schulte-Fortkamp *et al.* 2007). The meaning of sound as well as expectation, memory, state of mind all influences soundscape evaluation (Berglund *et al.* 1994; Yang and Kang 2005b; 2005a; Yu and Kang 2008; Lam *et al.* 2010). In addition to acoustical properties of sounds, non-acoustical features of the environment, including view, temperature, humidity, and wind, also have interactions with soundscape (Yang and Kang 2005a; Pheasant *et al.* 2008; Jeon *et al.* 2010a).

Although there are a number of influencing factors, almost all the relevant soundscape research show a similar tendency of human listeners of preferring natural sounds such as water sounds and birdsongs, rejecting mechanical sounds such as vehicles and construction sounds, and having neutral attitudes towards human sounds such as speech and footsteps (Yang and Kang 2005b; Nilsson and Berglund 2006; Axelsson *et al.* 2010). Moreover, benefits of natural sounds in improving people's mental health compared with noisy urban environmental sounds have been suggested, which include facilitating stress recovery and increasing memory retrieval (Alvarsson *et al.* 2010; Benfield *et al.* 2010).

However, reasons for such sound preferences still remain in question. We could ask, is natural sound good for physical health, or mental health, or makes people emotionally pleasant? And what properties of sound affects the preferences, the acoustic properties or information carried by sound? It is also expected that there are some interactions between these heath and emotion aspects. In order to answer the questions, to the first step, this research explores if there are any differences between natural and urban sounds in the aspect of acoustic properties. For next steps, if no differences existed, it could then conclude that other properties affect the preferences, e.g., the information. Conversely, if differences existed, how do these differences affect listeners, physiologically or psychologically? Answering these questions would need further studies of links between acoustic properties and health, or emotion.

This research thus works on the first step, and aims to study the possible differences and characteristics of various natural and urban environmental sounds in terms of objective measures. The next steps of research on sound preferences are outside the scope of this thesis. Although the differences between natural and urban sounds in influencing people's perception may have different facets, including acoustical, physiological, psychological and social, this study attempts to study the differences in a number of different factors in acoustic and psychoacoustic properties. These factors are chosen for having the potential to provide further links between acoustic properties and physiological or psychological facets. While it can be expected that the human hearing system has been adapted to common natural sounds, this research firstly seeks to study the differences from the aspect of subjective sensations of the human hearing system, in particular relating to the field of psychoacoustics. Psychoacoustics studies the quantitative correlation between acoustical stimuli and hearing sensations (Zwicker and Fastl 1999). Secondly, as soundscape and music are closely related (Schafer 1977), this study attempts to study the differences with music features, using the musical analysis techniques. In the field of music, the relation between musical features and humans' emotions and preferences has been researched for decades (Eerola *et al.* 2009; Han *et al.* 2009). It is

expected that these knowledge and techniques would benefit further study of the evaluation, both general and emotional effects, of soundscapes. Finally, it studies the differences in terms of 1/f noise dynamic. 1/f noise reflects the universal natural law of variation in nature and in soundscapes (Voss and Clarke 1978; De Coensel *et al.* 2003).

This study focus on sound types, thus only sound is concerned; the influence of environment and people in soundscapes are not considered, as the preference to natural sound is generally independent of these factors. Also, to simplify questions, only single source sounds are considered in this study. Mixture of various sound source components in soundscape environments, and the influences of environment to sound, e.g., reverberation (Kang 2001; 2002; 2005), are not considered.

1.2 Aim of the Study

The aim of this research is to study the characteristics of various natural and urban environmental sounds and examine their differences. It seeks to link the sound types with objective measures in three aspects: psychoacoustics, music features, and dynamic. In addition to the basic understanding of various sounds, the study of such differences would be useful for automatic identification and classification of sound types in soundscapes. The detailed research questions include:

1. Applicability of music features to environmental sounds and applicable algorithms of the music features;
2. Whether there are differences between natural and urban sound sources in terms of the three aspects, i.e., psychoacoustics, music features, and dynamic, and characteristics of various natural and urban environmental sounds and their differences if the differences existed;
3. The main factors that show differences between natural and urban sound sources if any; and
4. Methods for automatic identification and classification of sound types in soundscapes.

1.3 Thesis Outline

Chapter 1, “Introduction”, introduces the background and aim of this research. Chapter 2, “Literature review”, covers four research areas: soundscape, psychoacoustics, psychology of music, and 1/f noise dynamic. As sonic studies are being undertaken in many

independent but related areas, soundscape research acts as a middle ground to unify these researches. This research, in the area of soundscape, seeks to study natural sounds from three aspects: subjective hearing sensation that have been recommended in previous soundscape research, additional hearing sensation related to emotion, universal 1/f dynamic, which respectively are in relation to the remained three areas above. Chapter 3, "Methodology", describes the methodology used for sound recording collection and the analysis methods. Chapter 5, "Applicability of pitch algorithms to environmental sounds", and Chapter 6, "Applicability of rhythm algorithms to environmental sounds", based on the review of a number of music features and corresponding algorithms, study applicability of features and algorithms, and simplify, combine, modify or improve the algorithms for application in soundscape study.

Chapter 4, "Psychoacoustical evaluation of natural and urban sounds in soundscapes", Chapter 7, "Characteristics of natural and urban sounds in soundscapes in terms of pitch and rhythm features", and Chapter 8, "1/f noise behaviour of natural and urban sounds in soundscapes", discuss the differences between natural and urban sounds in three aspects: psychoacoustic parameters, music features, and 1/f noise dynamic. Respective characteristics of various natural and urban environmental sounds are shown, and models for automatic identification and classification of sound types are made. Chapter 9, "Characteristics of natural and urban sounds in soundscapes in terms of psychoacoustic, music parameters, and 1/f noise behaviour of the parameters", summarises the three aspects, and also factor analysis and identification are made based on all the aspects together. Chapter 10, "Conclusions and future works", summarises the contributions of the research. The outline of the thesis is shown in Figure 1.3.1.

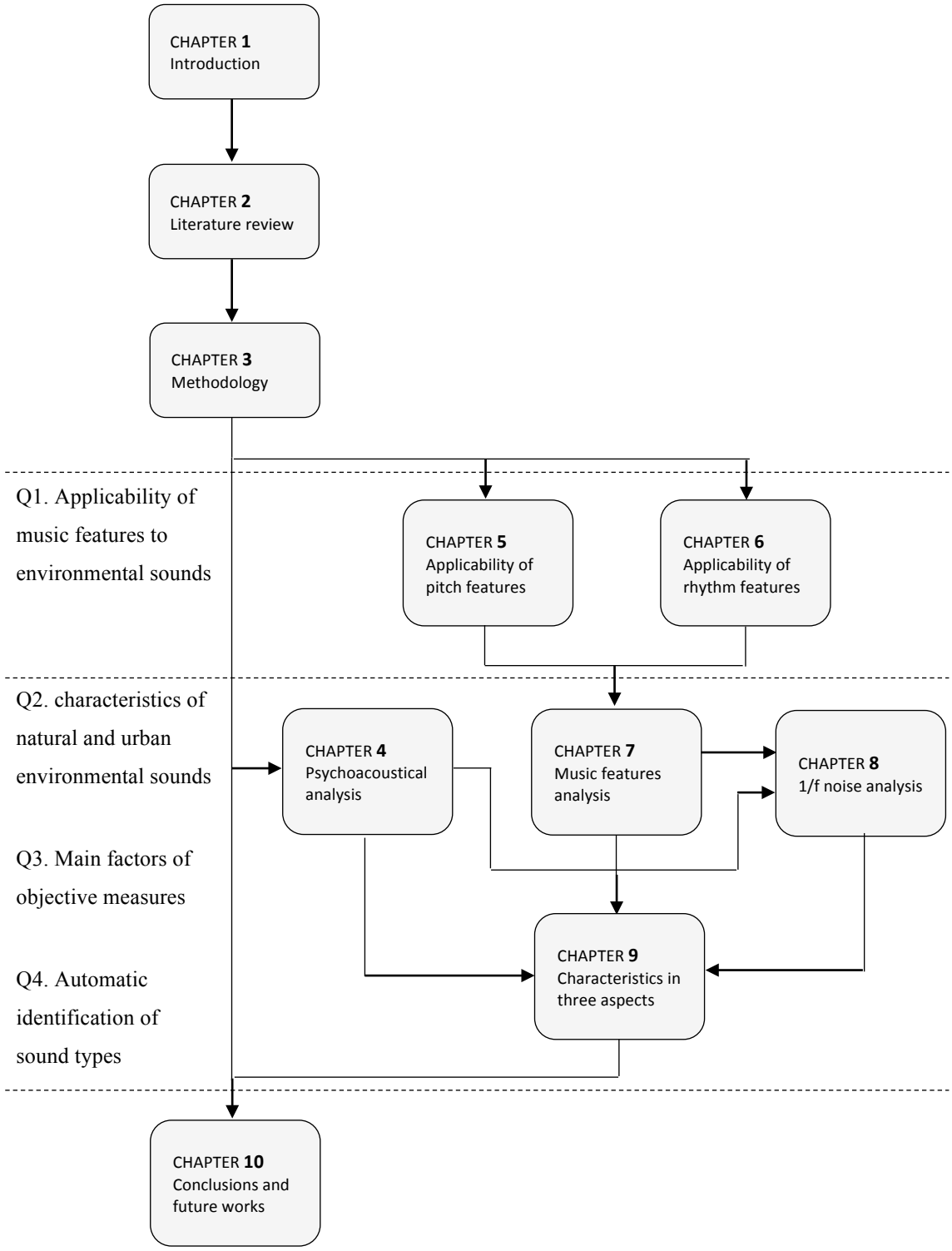


Figure 1.3.1 Thesis outline

Chapter 2
Literature review

As sonic studies are undertaken in many independent areas, soundscape studies attempt to unify these various researches, which are related and each deals with aspects of the soundscape. Soundscape studies would be "the middle ground between science, society and the arts" (Schafer 1977, p3-4). In Section 2.1, a framework of soundscape studies is established based on the elements in the process from sound being produced to being perceived by human. While soundscape studies covered all of these elements, this research concerns the sound source and feeling of human's ear. Consequently, three main areas of sonic studies are reviewed in this chapter, in addition to the area of soundscape.

Section 2.2 briefly reviews the area of soundscape, including the basic concept, research approaches, influences of the elements in the framework, i.e. sound, environment, people, to soundscape evaluation.

Section 2.3 reviews the scientific area of psychoacoustics, which concerns with the way that sound is perceived in the auditory system and the subjective sensations evoked by sound. A number of psychoacoustic parameters are reviewed, which have been applied to the fields of music and industrial acoustics design (Zwicker and Fastl 1999) and a part of which have been suggested in some soundscape studies (Genuit and Fiebig 2006).

Section 2.4 briefly reviews some parts of the area of psychology of music, which initially shares the same hearing part with psychoacoustics. It concentrates on cognition of musical sounds and the effects of music, more specifically, music and emotion. As music and soundscape are closely related (Schafer 1977), it is expected that these knowledge would benefit study of relationships between soundscape and emotion.

Section 2.5 reviews 1/f noise, which was found to be universal in nature, and was further measured in music, speech, and soundscapes (Voss and Clarke 1978; De Coensel *et al.* 2003). It reflects the natural law of variation in soundscapes. In Section 2.6, a summary of the reviews is made.

2.1 Introduction – A framework of soundscape

As soundscape concerns both the sonic environment and humans' perception (Schafer 1977; Truax 1999), it is reasonable to identify the elements or steps that are relevant to

the whole process from sound being produced to being perceived by human. From a physical perspective, the system of the process consists of a chain of elements: source, medium and receptor. In other words, there are first the source that emits sound, second the air that transmits and third the listener who detects. Energy transmits through the elements in the whole process, in one of its multiple forms (such as sound waves) depending on the particular step.

When looking closer at each step involved, several secondary components can be identified (Roederer 1995). At the source, there are two distinct components: the primary *excitation mechanism* (or force) that acts as the primary energy source, and the *vibrating object* that produce sound, when excited by the primary mechanism. This vibrating object actually determines the spectrum (pitch and timbre) of the sound. It converts the mechanical vibration of object into sound – oscillation of the surrounding air. In the medium, there are the *medium material* (air in most soundscapes) through which the sound propagates, and the *boundaries*, that is, the space environment such as surrounding buildings in urban place, which affect the sound propagation by reflection and absorption of the sound waves and whose configuration determines the quality of reverberation. Finally, for the listener, the components are the *peripheral auditory system* which includes eardrum, that picks up the pressure oscillations of the sound wave reaching the ear and transforms them into mechanical vibrations, and the inner ear, or cochlea, in which the vibrations are converted into electrical nerve impulses; and the *auditory neural system*, which transmits the neural signals to the brain where the information is processed, leading to the perception and cognition of sounds.

Consequently, changes of any of the components from sound being produced to perceived by human, i.e. force that causes object's vibration, vibrating object, medium through which sound propagates, environment that effects sound propagation, perception of the human auditory system, and cognition of higher level neural system of brain, would change the whole process in some specific way, and thus influence the final humans' perception and evaluation of soundscape.

A framework of soundscape can be thus established on the basis of these three main elements or aspects, which are sound, environment and people, as well as the final evaluation of soundscape. Here, in addition to the physical concepts of source, medium and receptor, people's perception is also involved, e.g., sound includes the meaning of sound, environment includes context and the vision aspect, and people includes memories and social factors. The main structure of the framework is illustrated with an example shown in Figure 2.1.1, based on various studies of soundscape. The studies have mainly focused on one particular aspect or the relationships between any two or more of them, illustrated with connection lines in the figure for a number of the studies.

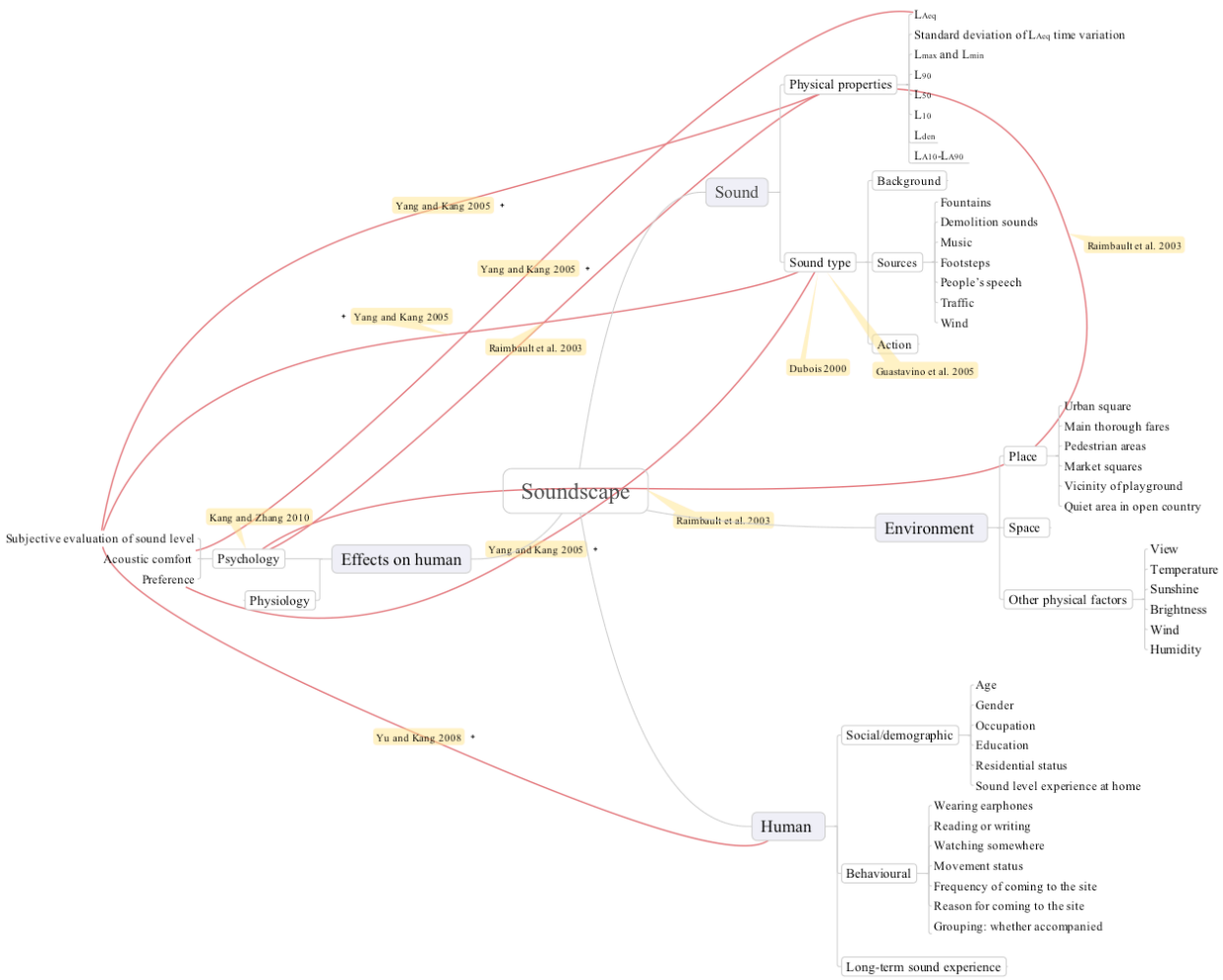


Figure 2.1.1 An illustration of the framework of soundscape

This research focuses on the relationships between sound sources and subjective auditory sensations, which respond to the elements/components of sound and perception of auditory system among those in the whole process of sound from being produced to perceived. Accordingly, in addition to the area of soundscape which attempts to join together researches in the various areas, three independent areas of sonic studies are reviewed in this chapter, i.e., psychoacoustics, psychology of music, and 1/f noise dynamic, among which psychoacoustics corresponds to the auditory system that includes both the peripheral and neural system, and psychology of music corresponds to cognition of neural system of brain.

2.2 Soundscape

This section briefly reviews the area of soundscape, beginning with basic concept and research approaches, and following by the respective discussions of each of the elements or aspects, i.e. sound, environment, people, and soundscape evaluation. Finally, the connections that emphasises people's preferences in terms of sound types are reviewed.

2.2.1 Soundscape concept

In traditional environmental noise control, considerable efforts have been made to reduce sound level, but it is gradually noticed that this does not necessarily lead to a better acoustic comfort (Kang 2006). Soundscape, a more positive way to deal with the environment noise pollution problem, has got much attention in the last few years. It was initially proposed by R. M. Schafer (1977), a Canadian musician, through the work of the World Soundscape Project (WSP) during the late 1960s and early 1970s (Truax 1999). Different from traditional noise control, soundscape concerns not only the sonic environment but also the perception of humans, as well as changes in the soundscape over time and across cultures (Schafer 1977; Truax 1999).

There is no single agreed definition of soundscapes (Cain et al. 2013). Schafer (1977) defined a soundscape as "the total acoustic environment", while Truax (1999) defined it as an environment of sound where the emphasis is on the way the sound is perceived and understood by an individual, or by a society. Thompson (2002) defined the soundscape as an auditory or aural landscape. Like a landscape, a soundscape is simultaneously a physical environment and a way of perceiving that environment (Davies et al. 2007).

Soundscape attempts to draw the independent areas of sonic studies together to answer the shared question that "the relationship between man and the sounds of his environment" (Schafer 1977, p3). These different areas include acoustics, psychoacoustics, otology, medicine, anthropology, ethnology, psychology, philosophy, communication, information, linguistics, media arts, musicology, literature, aesthetics, architecture, landscape, design, urban planning, ecology, noise control engineering, technology, sociology, religious studies, political science, law, pedagogics, and etc. (Schafer 1977; Karlsson 2000; Kang 2006). As a multidisciplinary, thus, soundscape brings together researchers from a wide range of academic backgrounds and have exchanges between the different disciplines.

2.2.2 Soundscape research approach

A large number of soundscape researches focused on the relationships of perception and evaluation of soundscape with the particular aspect of sound, environment or people, or with more than one aspect of them. The related soundscape research approaches thus combine measurements of sound, environment and people with scientific investigations of people's perception of soundscape (Schulte-Fortkamp *et al.* 2007).

The initial approaches involve soundwalks (Schafer 1977; Berglund and Nilsson 2006; Jeon *et al.* 2010b), which generally refer to route walks when listening and experiencing the sonic environment. It enables the access of perceptions of soundscape in all aspects of sound, environment (context and views etc.) and people. A large number of soundscape studies utilised field surveys, in which physical measurements of sound and environment and perceptual interviews or questionnaires were conducted simultaneously, together with demographic data of interviewees (Yang and Kang 2005a). It ensures the conservation of interactions between sound and other environment and social factors.

For analysis of long-term soundscapes, open, narrative and issue-centred interviews were usually carried out, while data of physical measurements might be obtained through local police or other public workers (Schulte-Fortkamp *et al.* 2007).

For more controlled conditions of research, recorded sound could be reproduced in laboratory and perceptual descriptions of soundscapes were collected through listening tests. Guastavino et al. (2005) investigated ecological validity of soundscape reproduction system in laboratory conditions. The experiments contrasted the verbal data of spontaneous descriptions of soundscapes in a field survey and two listening tests, using spatially different reproduction schemes in an acoustically damped room. The first listening test used stereophonic reproduction with two speakers, and the second used

multichannel ambisonics reproduction. Both laboratory conditions were found to be ecologically valid in terms of source event identification. However, regarding background noise, in the first laboratory condition, it was primarily described in terms of physical properties, rather than subjective effects, as in the field survey. In the second listening test, descriptions collected were similar to the field survey. These results confirmed the ecological validity of the ambisonics multichannel reproduction with no visual reference to the speakers, while the stereophonic reproduction did not enable the ecological validity for subjective effects of background noise. Also, 2-D configurations of spatial reproduction systems were preferred for perceptual evaluation of recreation of outdoor environments, whereas 3-D configurations appeared to be more preferred for the recreation of indoor environments (Guastavino and Katz 2004).

2.2.3 Soundscape evaluation

The assessment or evaluation of soundscapes may be accessed through scientifically developed interviews, questionnaires or other means. As a scientific tool, open interviews or questionnaires may yield more valuable information, while structured questionnaires, such as using semantic differential, are easier to conduct and measure.

The subjective evaluation of soundscapes often involves descriptors of emotional reaction and feeling related to the acoustic field. The relevant studies revealed principle components in the assessment of soundscapes (Axelsson *et al.* 2010; Kang and Zhang 2010; Cain *et al.* 2013). Although the descriptors differ between studies, some broad lines emerge (De Coensel and Botteldooren 2006). A first factor, which may be considered as the most important factor in most studies, is mainly associated with pleasantness, comfort or loudness of soundscape (Axelsson *et al.* 2010; Kang and Zhang 2010). A second factor is generally associated with dynamic, eventfulness or activity (Axelsson *et al.* 2010; Cain *et al.* 2013). The next factors may be associated with spatiality (Raimbault *et al.* 2003; Kang and Zhang 2010), familiarity (Axelsson *et al.* 2010), or timbre of soundscapes (Zeitler and Hellbrück 2001).

2.2.4 Sound factors

Two types of listening processing regarding soundscapes were identified, which aimed either at identification of acoustic sources and events or at qualification of soundscape as a whole (Dubois *et al.* 2006). From the linguistic analysis of subjects' perceptual descriptions of soundscapes, two broad categories were derived: source events, and

background noise of environment, where no specific events could be discriminated. Source events were primarily described with reference to specific sources or objects producing noise, whereas background noise was described both in terms of physical properties of the acoustic signal and of its effects on the subjects (the same as in the case of the source events) (Guastavino *et al.* 2005).

The indicators of physical properties ranged from measurement of overall A-weighted sound pressure level (SPL) and spectral content, through psychoacoustic measures, such as loudness, sharpness (essentially the ratio of high-frequency loudness to overall loudness), roughness (quick fluctuation) and fluctuation strength (slow fluctuation) (also see Section 2.3), and further time histories and statistical analyses of time-variance of these measures (Schulte-Fortkamp *et al.* 2007). The physical indicators could generally be represented with a number of factors (De Coensel and Botteldooren 2006). Based on the research results of relevant studies, the sound strength was recognized as an important factor, such as L_{Aeq}, L_{A10}, N_{10} (Raimbault *et al.* 2003), statistical levels between L_{A50} and L_{A95} for quiet areas (De Coensel and Botteldooren 2006), and especially the background sound level (L_{A90}), for which it was found that a lower background sound level could make people feel quieter (Yang and Kang 2005a).

A second factor of physical indicators could refer to the dynamics and temporal structure of soundscape, e.g. L_{A10}-L_{A90} (Nilsson 2007) and indicators of 1/f noise behaviour (Voss and Clarke 1978; De Coensel *et al.* 2003). The 1/f noise indicators were calculated from the spectrum (e.g., slope and deviation from a straight line) of the temporal envelope (short term loudness or L_{Aeq}) of the sound and represented the temporal structure of soundscape (Licitra *et al.* 2005; Botteldooren *et al.* 2006). 1/f noise, which is found to be universal in nature and reflects the natural law of variation, has been measured in music, speech and soundscape (Voss and Clarke 1978; De Coensel *et al.* 2003) (also see Section 2.5 and Chapter 3 Section 3.3.3). A third factor was related to the spectral content of sound, such as the centre of gravity of the spectrum (Raimbault *et al.* 2003), or related to distance perception (Preis and Golebiewski 2004).

2.2.5 Environment factors

The influence of environmental factors to the acoustic comfort evaluation has been studied in 14 urban open public spaces across five European countries (Yang and Kang 2005a; Yu and Kang 2009), in terms of various physical indices including view, temperature, brightness, wind and humidity. It was found that the acoustic comfort evaluation was statistically significantly related to the subjective evaluations of view and

brightness, while the correlations to the evaluations of temperature, wind, and humidity were rather limited. Also, principal component analysis of the physical indices showed that visual and auditory aspects were always in the same factor, suggesting that these two aspects might have certain interactions (Yang and Kang 2005a).

The interaction between visual and aural stimuli on the perceptions of soundscape or of total environment has been researched in a number of studies. In Southworth's (1969) research, by analysing the reports of visual and auditory elements saw or heard of three groups of subjects, including auditory only subjects, visual only subjects and visual-auditory subjects, during a tour in Boston city, it was found that when the sounds are related to the scenes, the interactions between visual and auditory perception increase sense of involvement. Carles et al. (1999) carried out research in laboratory condition with sounds and images of natural and semi-natural settings and urban green spaces. Subjects rated the auditory and visual stimuli separately and in varying combinations in terms of pleasure. The results showed that sound provides additional information to visual data and thus influences assessment of total landscape. Also with controlled auditory and visual stimuli in laboratory conditions, Viollon's (2003) research found that for the overall urban sound scenes, the more pleasant the visual setting, the less contaminated the auditory judgement. Pheasant et al. (2008) studied the influence of combined audio-visual modality to the perception of tranquillity of open spaces. The results showed that the subjective assessment of tranquillity of bi-modal (audio-visual) data approximated a mid-way value between those of the audio and video only. Both the audio factor of maximum sound pressure level (L_{Amax}) and the visual factor of percentage of natural features present at a location were key in influencing the perception of tranquillity.

2.2.6 People factors

Behavioural and social/demographic factors may also influence the evaluation of soundscapes. Based on large scale field surveys in urban open spaces across Europe, Yang and Kang (2005a) studied the effects of demographic factors on acoustic comfort evaluation, such as age, gender, occupation, and education. The results showed that no significant difference was found between males and females, local and non-local people or different occupations, other than different age groups. Teenagers tended to be most unsatisfied, whereas older people were most satisfied. Differences have also been found among different age groups in terms of sound preference (discussed below) (Yang and Kang 2005b).

2.2.7 Sound preferences

The preference of sound is an important aspect in soundscape evaluation. In a field study where subjects' preference of sound during a tour in central Boston was analysed, Southworth (1969) found that pleasantness of sounds appeared to depend on much more than the physical qualities of the sound. *"Low to middle frequency and intensity sounds were preferred, but delight increased when sounds were novel, informative, responsive to personal action, and culturally approved"* (Southworth 1969).

In the World Soundscape Project (Schafer 1977), it tested the subjects' most favourite and least favourite sounds in many different countries, of which the general patterns produced supported that different cultural groups have varying attitudes to environmental sounds. Climate and geography influence likes and dislikes to some extent. For instance, *"while in countries which touch the sea, ocean waves are well liked, in an inland country like Switzerland, the sounds of brooks and waterfalls are a much greater favourite"* (Schafer 1977, p159). Also, reactions to sound are affected by the degree of proximity to the elements. Technological sounds, such as machine sounds and traffic noise, are strongly disliked in all technologically advanced countries, while they may be liked in parts of the world where they are more novel. This is in correspondence with the findings by Southworth (1969) that delight increased when sounds were novel.

In Carles et al.'s (1999) research in laboratory condition, subjects' response to sounds and images of natural/semi-natural settings and urban green spaces was measured in terms of pleasure. The results showed a rank of preferences running from natural to man-made sounds. Natural sounds, meanwhile, might improve the quality of built-up environments to a certain extent.

In a cross-cultural soundscape study (Yang and Kang 2003), over 6000 people were interviewed in 14 urban open public spaces of five European countries. The study suggested that although sound preferences may be significantly different due to differences in cultural background and long-term environmental experience, people from different countries showed a similar tendency of preferring nature and culture-related sounds and rejecting vehicle and construction sounds. Later, Yang and Kang (2005b) carried out an intensive field questionnaire survey in two typical urban squares in Sheffield over four seasons with over 1000 interviews. Similarly, the study found *"people showed a very positive attitude towards the natural sounds. More than 75% of the interviewees were favourable to water sound and birdsong, and only less than 10% of the people thought the sounds were annoying… For culturally approved sounds, such as church bells, music played on the street and clock chimes or music, people also showed relatively high levels of preference. For human sounds such as surrounding speech, most*

people thought they were neither favourite nor annoying. The most unpopular sounds were mechanical sounds, such as construction sounds, music from passenger cars and vehicle sounds…" In terms of the effects of demographic factors in soundscape preferences, it was shown that, generally speaking, with an increase in age, people are more favourable to, or tolerant towards, sounds relating to nature, culture or human activities. By contrast, younger people are more favourable to, or tolerant towards, music and mechanical sounds.

2.2.8 Soundscape identification

Since the recognition and subjective evaluation/preference of soundscapes operate on the basis of identification of physical sound sources as discussed above, numbers of studies aimed to build a system that can become the basis for an automatic soundscape analysis tool by identifying sound events in soundscapes.

For single environmental sound recognition, Cowling and Sitte (2003) comprehensively compared the different techniques, that were typically used in speech/speaker and musical instrument recognition, in their suitability for environmental sound identification. Basically, sound recognition (both speech and environmental) is achieved by two phases: first feature extraction, followed by classification. The feature extractor produces a set of characteristic features for sound to reduce the complexity of the data before it reaches the classifier. The classifier is then used to recognise the sound based on the extracted features. From the combinations of feature extraction techniques, such as frequency extraction, homomorphic/ Mel frequency/ Bark frequency cepstral coefficients, linear prediction cepstral (LPC) coefficients, Mel frequency/ Bark frequency LPC coefficients, perceptual linear prediction (PLP) features, short-time Fourier transform (STFT), fast (discrete) wavelet transform (FWT), continuous wavelet transform (CWT), and Wigner-Ville distribution (WVD), and classification techniques, such as dynamic time warping (DTW), hidden Markov models (HMM), learning vector quantization (LVQ), self-organising maps (SOM), artificial neural networks (ANN), long-term statistics, maximum likelihood estimation (MLE), Gaussian mixture models (GMM), and support vector machines (SVM), Cowling and Sitte (2003) found the combination of continuous wavelet transform or mel frequency cepstral coefficients (MFCCs) with dynamic time warping produced a classification rate of 70%. In the project of instrument for soundscape recognition, identification and evaluation (ISRIE), Bunting et al. (2009) used time-domain signal coding (TDSC) combined with LVQ network.

For real-world environmental sounds, ISRIE project (Bunting *et al.* 2009) employed a source separation algorithm prior to the recognition stage. Krijnders et al. (2010) also improved the signal-driven classification, performed by segment and feature extraction from a time–frequency cochleogram and machine-learning techniques, by creating expectancies of sound events based on context information through a dynamic network. In contrast to these methods, Aucouturier et al. (2007; Aucouturier and Defreville 2007) proposed to directly recognize soundscapes holistically, without the prior identification of constituent sound sources, using the "bag-of-frames" approach. It represented signals as the long-term statistical distribution of frame-based MFCC vectors, using GMMs, and proved precision of 0.9 in the first five nearest neighbours. Rychtáriková and Vermeir (2013) categorised sound recordings of urban public places based on acoustical multi-parameter analysis, using sound pressure level and psychoacoustical parameters. The objective clustering was found to be consistent with subjective expectations on the basis of the typology of the recording locations and activities.

2.2.9 Discussions

As reviewed above, with physical, linguistic, psychological, sociological, and other approaches in the soundscape studies, soundscape evaluations have been intensively studied (Raimbault and Dubois 2005; Kang 2006). These researches showed that soundscape evaluations are influenced by all factors of the aspects of sound, environment, and people, which cover the scope of soundscape (Botteldooren and Verkeyn 2002; Raimbault and Dubois 2005; Genuit and Fiebig 2006; Kang 2006; Schulte-Fortkamp and Fiebig 2006; Schulte-Fortkamp *et al.* 2007). For example, the meaning of sound as well as expectation, memory, state of mind all influences soundscape evaluation (Berglund *et al.* 1994; Yang and Kang 2005b; 2005a; Yu and Kang 2008; Lam *et al.* 2010). In addition to acoustical properties of sounds, non-acoustical features of the environment, including view, temperature, humidity, and wind, also have interactions with soundscape (Yang and Kang 2005a; Pheasant *et al.* 2008; Jeon *et al.* 2010a).

Although there are a number of influencing factors, almost all the relevant soundscape research show a similar tendency of human listeners of preferring natural sounds such as water sounds and birdsongs, rejecting mechanical sounds such as vehicles and construction sounds, and having neutral attitudes towards human sounds such as speech and footsteps (Yang and Kang 2005b; Nilsson and Berglund 2006; Axelsson *et al.* 2010). Moreover, benefits of natural sounds in improving people's mental health compared with noisy urban environmental sounds have been suggested, which include

facilitating stress recovery and increasing memory retrieval (Alvarsson *et al.* 2010; Benfield *et al.* 2010).

In this study, in order to explore the factors for such sound preferences, it intends to link the sound sources with objective measures of sound, which are related to subjective auditory sensations, since it can be expected that the human hearing system has been adapted to common natural sounds. In other words, this research focuses on the relationships between sound sources and perception of auditory system among those in the whole process of sound from being produced to being perceived. Here, the influences of environments and people to the evaluation are not considered.

Accordingly, three independent areas of sonic studies are reviewed following in this chapter, i.e., psychoacoustics, psychology of music and 1/f noise dynamic, in addition to the area of soundscape which attempts to join together researches in the various areas. Psychoacoustics concerns with the way that sound is perceived in the auditory system and the relationship between the objective, physical properties of acoustical stimuli and the subjective, psychological sensations evoked by them. Research in the psychology of music uses psychological theories and methods to interpret and understand musical behaviours, that psychology of music initially shares the same hearing part with psychoacoustics. It also concentrates on cognition of musical sounds and emotional effects of music. 1/f noise reflects the natural law that is found to be universal in the dynamic of noise, and in music, speech, and soundscapes.

2.3 Psychoacoustics

Psychoacoustics, a part of psychophysics, is a scientific field concerning with the relationship between the objective, physical properties of acoustical stimuli and the subjective, psychological responses evoked by them (Rasch and Plomp 1982, p1). Psychoacoustics as a scientific field has a tradition of more than 2500 years (Fastl 2005). Already around 500 B.C. the Greek philosopher Pythagoras had studied musical consonance and dissonance with a monochord using the method of psycho-acoustical experiments (Fastl 2005).

The quantitative correlation of acoustical stimuli and hearing sensations, or say response characteristics, is investigated through the experiment in laboratory (Zwicker and Fastl 1999, pVII), which determines psychoacoustics as an empirical, or experimental science (Rasch and Plomp 1982, p2). A psycho-acoustical experiment can be described most simply in a stimulus-response scheme. The stimulus is the sound presented to the subject. The subject is required to give a response (Rasch and Plomp 1982, p2). Psycho-

acoustical research investigates the correlation through experimental data and models the measured facts in an understandable way (Zwicker and Fastl 1999, pVII), often without explaining the relationships in terms of the underlying mechanisms of sensory processes (Rasch and Plomp 1982, p2). While some aspects of auditory perception can be explained by reference to studies in physiological acoustics, our knowledge in this respect so far is usually limited to possibly specify the detailed physiological mechanisms (Rasch and Plomp 1982, p2; Moore 1997, p1).

In this section, the basic physiology of the auditory system and its frequency-analysing power, a capacity that is fundamental to its perceptual functioning, are firstly described. Then, the aspects of subjective properties, i.e. loudness, pitch, timbre, etc., and algorithms for simulating these auditory sensations are briefly reviewed.

2.3.1 Auditory system

Stimulus processing in the auditory system consists of pre-processing of sound in the peripheral system and information processing in the neural system (Zwicker and Fastl 1999, p23). In the peripheral auditory system, which comprises the outer, middle, and inner ear (Moore 1997, p47), sound vibration patterns retain in the form of mechanical oscillations. The sensory cells receive pre-processed oscillations from the peripheral structures, and encode the mechanical stimuli into electrical action potentials through nerve terminals in the neural system. The neural processing finally leads to auditory sensations (Zwicker and Fastl 1999, p23).

2.3.1.1 Basic structure of and preprocessing of sound in the peripheral system

Sound travels through the outer ear, composed of the pinna and the outer ear canal, and causes the eardrum to vibrate (Moore 1997, p17). The outer ear, together with other parts of body, such as the head, distorts the sound field and influences the sound coming to the eardrum (Zwicker and Fastl 1999, p23). Shoulders and head influence the sound pressure level in front of the eardrum *"most effectively at frequencies below 1500Hz through shadowing and reflection"* (Zwicker and Fastl 1999, p23). The outer ear canal *"acts like an open pipe with a length of about 2 cm corresponding to a quarter of the wavelength of frequencies near 4 kHz"*. It is *"responsible for the high sensitivity of our hearing organ in this frequency range, indicated by the dip of threshold in quiet around 4 kHz"* (Zwicker and Fastl 1999, p24).

The major function of the middle ear is to efficiently transfer sound, the oscillations of air particles from the outer ear, into motions of the salt water-like fluids in the inner ear

(cochlea) (Zwicker and Fastl 1999, p24-25). Because the difference in acoustical impedance of the two fluids - air outside and water inside, the middle ear acts as an impedance-matching device or transformer that improves sound transmission and reduces the amount of energy losses through reflections (Moore 1997, p18; Zwicker and Fastl 1999, p25). This is accomplished by the difference in areas of the eardrum and the oval window, the ratio of which is about 15, and by the lever action of the ossicles - a lever ratio of about 2 produced by the different lengths of the arms of the malleus and incus (Moore 1997, p18; Zwicker and Fastl 1999, p25). Through the area and lever ratios, an almost perfect match between the impedances is reached in the frequency range around 1 kHz (Zwicker and Fastl 1999, p25), and transmission of sound through the middle ear is most efficient at middle frequencies of 500-4000 Hz (Moore 1997, p18).

The inner ear (cochlea) is the most important part of the ear for understanding many aspects of auditory perception (Moore 1997, p19). The cochlea, shaped like a snail, is filled with almost incompressible fluids and has bony rigid walls (Moore 1997, p19; Zwicker and Fastl 1999, p25). It is divided along its length - from the base to the apex - by two membranes, Reissner's membrane and the basilar membrane (BM), into three channels or scalae (Moore 1997, p19; Zwicker and Fastl 1999, p25). The cochlea forms 2 1/2 turns allowing a basilar membrane length of about 32mm (Zwicker and Fastl 1999, p25).

"The inner ear performs the very important task of frequency separation: energy from different frequencies is transferred to and concentrated at different places along the basilar membrane" (Zwicker and Fastl 1999, p29). The hypothesis of von Helmholtz (von Helmholtz 1863; von Bekesy 1960), launched about 150 years ago, that the cochlea performs a frequency analysis which sound components with high frequencies produce oscillations of the basilar membrane close to the base (oval window) and components with low frequencies near the apex (helicotrema), was confirmed by experimental results of von Békésy (von Bekesy 1947; 1960; Rasch and Plomp 1982, p4; Zwicker and Fastl 1999, p28).

Von Békésy (1947; 1960) discovered the existence of travelling waves in contract to the previously conceived standing waves. When an incoming sinusoidal stimulation sets the oval window in motion and thus causes the BM to move, the vertical displacement of the BM takes the form of a travelling wave which moves along the BM from the base towards the apex. The amplitude of the wave is small at first, increases gradually, reaches a maximum at a certain location, and then decreases rather abruptly (Zwicker and Fastl 1999, p28-29). For each frequency there is a maximum in the pattern of vibration at a different place along the BM. This response of the BM to different frequencies is strongly affected by its mechanical properties, which vary considerably from base to apex. At the

base it is relatively narrow and stiff, while at the apex it is about three times wider and much less stiff (Moore 1997, p19). *"As a result, the position of the peak in the pattern of vibration differs according to the frequency of stimulation. High-frequency sounds produce a maximum displacement of the BM near the oval window, with little movement on the remainder of the membrane. Low-frequency sounds produce a pattern of vibration which extends all the way along the BM, but which reaches a maximum before the end of the membrane."* (Moore 1997, p19-20).

As sounds of different frequencies produce maximum displacement at different places along the BM, the cochlea thus behaves like a Fourier analyser, although with a limited frequency-analysing power. The frequency scale of sound is converted into a spatial scale along the BM (Rasch and Plomp 1982, p4).

The mechanical oscillations of the BM is converted or coded into neural signals in the auditory nervous system through the organ of Corti. It contains hair cells, which are between the BM and the tectorial membrane. The hair cells are divided into two groups - inner hair cells and outer hair cells - by the tunnel of Corti (Moore 1997, p28). The tectorial membrane lies above the hairs of the hair cells (Moore 1997, p28). *"It appears that the hairs of the outer hair cells actually make contact with the tectorial membrane, but this may not be true for the inner hair cells"* (Moore 1997, p28). Movement of the BM causes a displacement of the hairs at the tops of the hair cells by the shearing motion between the hairs and the tectorial membrane.

The construction of inner and outer hair cells is different, which indicates the different functions (Zwicker and Fastl 1999, p27-28). The great majority (more than 90%) of the afferent neurones connect to inner hair cells. Thus, most information about sounds is conveyed via the inner hair cells (Moore 1997, p29). The outer hair cells make contact with many efferent fibres coming from higher centres of the auditory system, which can thus affect their activity. There is also evidence that *"the outer hair cells have a motor function, changing their length and shape in response to electrical stimulation"* (Ashmore 1987; Moore 1997, p29). These support the idea that the outer hair cells play an active role in influencing the mechanics of the cochlea, so as to produce high sensitivity and sharp tuning (Moore 1997, p29).

2.3.1.2 Neural responses in the auditory nerve

Activity in the auditory nerve was studied mostly by recording the nerve impulses, or spikes, in single auditory nerve fibres (Moore 1997, p31). Deflection of the hairs on the inner hair cells produced by movement of the BM towards the tectorial membrane leads to neural excitation, or firing in nerve fibre. No excitation occurs when the BM moves in the opposite direction (Moore 1997, p39).

The nerve fibres show spontaneous firing in the absence of sound stimulation (Moore 1997, p31). On the basis of the spontaneous rates, auditory nerve fibres could be classified into three groups: *"about 61% of fibres have high spontaneous rates (18 to 250 spikes per second); 23% have medium rates (0.5 to 18 spikes per second); and 16% have low spontaneous rates (less than 0.5 spike per second)"* (Liberman 1978; Moore 1997, p32).

When a sound stimulation is presented, above a threshold level, the firing rate of an auditory nerve fibre increases with an increase of stimulus level. The threshold of a neurone is the lowest sound level at which a change in firing rate can be detected. Above a certain sound level, say a saturation level, the neurone no longer responds to increases in sound level with an increase in firing rate. The range of sound levels between threshold and saturation is called the dynamic range. High spontaneous rates tend to be associated with low thresholds and small dynamic ranges (Moore 1997, p34-35).

The nerve fibres respond better to some frequencies than to others; they show frequency selectivity (Moore 1997, p31). The threshold of a given fibre is lowest for one frequency, called the characteristic frequency (CF) and increases for frequencies on either side of this (Moore 1997, p47). It is generally assumed that the frequency selectivity in single auditory nerve fibres occurs because a single fibre is responding to activity at a particular part of the BM (Moore 1997, p32).

Besides the rate of firing, the temporal pattern of firing also carries information about the stimulus. Neural firings in the fibres tend to be phase locked or synchronized to the stimulating waveform, that is to occur at a particular phase of the waveform (Moore 1997, p38,47). *"A given nerve fibre does not fire on every cycle of the stimulus but, when firings do occur, they occur at roughly the same phase of the waveform each time. Thus the time intervals between firings are (approximately) integral multiples of the period of the stimulating waveform"* (Moore 1997, p38). Thus, the distribution of time intervals between successive nerve firings, or the temporal pattern of firing responds to the frequency of the stimulating waveform. Phase locking breaks down about 4-5 kHz (Palmer and Russell 1986).

2.3.2 Critical bands

The frequency selectivity of the hearing system as described above, indicates that it can be assumed that the hearing system processes sounds in relatively narrow frequency bands (Zwicker and Fastl 1999, p149), and thus, the auditory system has often been modelled as a bank of overlapping bandpass filters, known as auditory filters (Moore

1997). The concept of critical bands was proposed by Fletcher (1940). He assumed that the part of a noise that is effective in masking a test tone is the part of its spectrum lying near the tone; parts of the noise outside the spectrum near the test tone do not contribute to masking (Zwicker and Fastl 1999, p149).

Zwicker and Fastl (1999), from the average of data using five methods, produced a reasonable estimation of the critical bandwidth (CBW). At low frequencies, critical bands show a constant width of about 100 Hz, while at frequencies above 500 Hz critical bands show a bandwidth which is about 20% of centre frequency. The audible frequency range (up to 16 kHz) is accordingly subdivided into 24 abutting critical bands, with a unit of "Bark", leading to the so-called critical-band rate scale or Bark scale.

In addition to Bark scale, Glasberg and Moore (1990) proposed equivalent rectangular bandwidth (ERB) scale, measured using the notched-noise method. The ERB of the auditory filter is assumed to be closely related to the critical bandwidth, and is generally narrower than the Bark scale CBW, being about 25 Hz at low frequencies and about 11% of center frequency at high frequencies.

The critical-band rate is closely related to several other scales that describe characteristics of the hearing system, such as the just-noticeable increment in frequency and ratio pitch, and to the function relating frequency to the position of maximal stimulation on the basilar membrane (BM). The width of the critical band corresponds to a distance along the BM of about 1.3 mm. *"Assuming that the abutting haircells have a distance of about 9μm along the whole length of the 32 mm basilar membrane, the total number of 3600 haircells in one row from helicotrema to oval window is achieved"* (Zwicker and Fastl 1999, p161).

2.3.3 Loudness

Loudness is defined as that attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud (Fletcher and Munson 1933; Moore 1997, p49). It is the sensation that corresponds most closely to the sound intensity of stimulus (Fletcher 1934; Zwicker and Fastl 1999, p205). Loudness depends on frequency and intensity (Moore 1997, p49), and also on many more variables such as bandwidth, frequency content, and duration (Zwicker 1965; Zwicker and Fastl 1999, p205). As a subjective quantity and, as such, cannot be measured directly, loudness has been studied in a number of ways of such as loudness comparison and magnitude estimation (Zwicker 1966; Moore 1997, p49; Zwicker and Fastl 1999, p203).

The dependence of loudness on frequency can be indicated by equal-loudness contours, which are generated by equal loudness levels of sinusoids of different frequencies (Fletcher and Munson 1933; ISO 226 2003). Loudness level of a sound is the sound pressure level of a 1000-Hz tone that is as loud as the sound, i.e. gives equal loudness. It is measured by loudness comparison, in the unit of 'phon' (Moore 1997, p54; Zwicker and Fastl 1999, p203). Curve of absolute threshold, or threshold in quiet, i.e. the minimum detectable level of a sound in the absence of any other external sounds (Moore 1997, p49), is also an equal-loudness contour indicated by 3 phon, which corresponds to threshold in quiet of 3 dB at 1kHz (Zwicker and Fastl 1999, p203). Absolute thresholds increase rapidly at very high and very low frequencies (Moore 1997, p51), which means that we are most sensitive to middle frequencies (1-5 kHz) (Moore 1997, p85). The equal-loudness contours are of similar shape to the threshold curve, but tend to become flatter at high loudness levels (Moore 1997, p54), that is at high levels tones of equal SPL sound roughly equally loud regardless of frequency (Moore 1997, p85). The shapes of equal-loudness contours indicate the dependence of loudness on frequency (Zwicker and Fastl 1999, p205), and also that the rate of growth of loudness level with increasing intensity is greater for low and very high frequencies than for middle frequencies (Moore 1997, p54).

Scales relating the physical magnitudes of sounds to their subjective loudness have been commonly derived by two methods, magnitude estimation and magnitude production (Stevens 1957; 1972; Moore 1997, p57). Average of many measurements of these kinds indicates that *"the loudness of a given sound is proportional to its intensity raised to the power 0.3". "A simple approximation to this is that a two-fold change in loudness is produced by a 10-dB change in level"* (Moore 1997, p58). With 'sone' as the unit of loudness (Stevens 1957; 1972), 1 sone is defined arbitrarily as the loudness of a 1-kHz tone at 40 dB SPL. A 1-kHz tone with a level of 50 dB SPL is usually judged as twice as loud as a 40-dB tone and has a loudness of 2 sones. At low levels, below 40 dB, this relationship does not hold, and the loudness changes more rapidly with sound level (Moore 1997, p58-59).

Models for estimating the loudness of sounds, incorporating the basic concept that loudness may depend upon a summation of neural activity across different frequency channels, have been proposed by Fletcher and Munson (1937), by Zwicker (Zwicker 1958; Zwicker and Scharf 1965) and by Moore and Glasberg (1996). Zwicker's loudness model has been standardized in international standards ISO 532B (1975). Essentially, there are three stages that form the basic of Zwicker's model (Zwicker *et al.* 1984; Fastl 2005). The first stage is the physical representation of sound along psychoacoustic critical-band scale. The second stage is the calculation of an excitation pattern accounting masking effects and threshold in quiet. This pattern can be thought of as representing the

excitation distribution at different points along the basilar membrane (BM). The third stage is the transformation from excitation pattern (level in dB) to loudness pattern, i.e. specific loudness as a function of critical-band rate. This transformation involves a compressive nonlinearity, using Stevens' power law and logarithmic transmission factor. This transformation can be thought of as representing the way that physical excitation is transformed into neural activity. The overall loudness of a given sound, in sones, is proportional to the total area under the specific loudness pattern (Moore 1997, p60).

2.3.4 Pitch

Pitch may be defined as "that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale" (ASA 1960; Moore 1997, p177). The pitch value to a sound is generally assigned by the frequency of a pure tone having the same subjective pitch as the sound (Moore 1997, p177). For a pure tone pitch is related to the frequency and for a periodic complex tone to the fundamental frequency (Rasch and Plomp 1982) p6, though there are exceptions to this simple rule (Moore 1997, p177). Pitch relationship is indicated by harmony when tones are presented simultaneously, and by melody when tones are presented sequentially (Burns and Ward 1982, p243). In this section the pitch of pure tones, complex tones and noise bands and corresponding pitch strength are first assessed, and then two main classes of pitch perception theories are described.

2.3.4.1 Pitch and pitch strength

For the pitch of pure tones, measurements show that at low frequencies, the halving of the pitch sensation corresponds to a ratio of 2:1 in frequency, at high frequencies above 1kHz, however, a ratio of larger than 2:1 is necessary for the perception of half pitch (Zwicker and Fastl 1999, p111). For example a 1300 Hz tone represents half pitch of that of 8 kHz. The pitch of a pure tone is primarily determined by its frequency, but also somewhat influenced by sound pressure level. *"On average, the pitch of tones below about 2kHz decreases with increasing level, while the pitch of tones above about 4kHz increases with increasing level"* (Moore 1997, p186). *"For an increase in sound pressure level of 40dB, the pitch of pure tones is shifted on average by not more than about 3%. This relatively small effect can be neglected in many cases"* (Zwicker and Fastl 1999, p114).

Complex tones can be regarded as the sum of several pure tones. The pitch of harmonic complex tones, for which the frequencies of the pure tone components are integer multiples of a common basic or fundamental frequency, basically corresponds to

the fundamental frequency (Zwicker and Fastl 1999, p118). If the lower harmonics are removed from a harmonic complex tone, the pitch hardly changes. This means that the pitch of incomplete harmonic tone without fundamental frequency, or say "residual" higher harmonics of a complex tone, usually corresponds closely to the pitch of the low or fundamental frequency. The missing fundamental effect has been termed residue pitch, low pitch, or virtual pitch (Zwicker and Fastl 1999, p120).

Noise with steep spectral slopes (at least 120 dB/octave) produces pitches that correspond to the frequencies of the spectral edges. For low-pass and high-pass noise, the pitch corresponds closely to the cut-off frequency of the filter (Zwicker and Fastl 1999, p124). Band-pass noise produces two pitches corresponding to the upper and lower cut-offs (Zwicker and Fastl 1999, p125). *"If the spectral edges are close together, the two edge pitches fuse to a single pitch corresponding to the centre frequency of the narrow-band noise"* (Zwicker and Fastl 1999, p127).

Experiments on pitch sensation described so far generally explore it along a scale from low to high, called pitch. Independent of the pitch, the sensation which can be labelled as faint pitch or distinct pitch leads to a scale of pitch strength (Zwicker and Fastl 1999, p134). The pitch strength of a variety of sounds described above, which include pure tones, complex tones, narrow-band noises, low-pass and high-pass noises, and band-pass noises, was measured using magnitude estimation. While all sounds can elicit approximately the same pitch, they differ considerably in pitch strength. Generally, sounds with line spectra elicit relatively large pitch strength, whereas sounds with continuous spectra produce only small pitch strength. The pitch strength of pure tones is the largest among all sounds. Complex tones on average produce at least half the pitch strength of a pure tone. The pitch strength of narrow-band noise is comparable to that of complex tones; and the pitch strength elicited by other types of noises is generally one fifth to one tenth of that of a pure tone (Fastl and Stoll 1979; Zwicker and Fastl 1999, p134).

Pitch strength of pure tones shows a dependence on frequency; it reaches largest values at mid frequencies (around 1.5 kHz). The pitch strength of pure tones increases with increasing duration, almost linearly with the logarithm of duration up to about 300ms, and with increasing sound pressure level - within a level range of 20 to 80 dB pitch strength increases by a factor of about 2.5. For band-pass noises, pitch strength has a dependence on bandwidth; it decreases with increasing bandwidth. Noises with steep spectral slopes produce pitch strength dependent on the steepness of the filter slope, or exactly the slope of the masking pattern - pitch strength increases almost linearly with the slope of its masking pattern. The pitch strength of pure tones can be reduced considerably by partial masking sounds. The pitch strength depends on the level of the tone above

masked threshold. The pitch strength of partially marked tones is very small when the level is only 3dB above masked threshold, and reaches almost half the value obtained with an unmasked pure tone for tones 10 dB above, and is almost equal to the pitch strength of an unaffected pure tone at levels of 20 dB above (Zwicker and Fastl 1999, p136-144).

2.3.4.2 Pitch perception theories

There have been two classes of theories of pitch perception regarding how the pitches of stimuli are related to the anatomical and physiological properties of the auditory system, besides to their physical properties (Moore 1997, p211). One of them is the 'place' theories, which proposes that the frequency of a sound may be coded by the distribution of activity across different auditory neurones. Different frequencies of a stimulus excite different places along the basilar membrane (BM) and hence neurones with different CFs, as the spectral analysis taking part in the inner ear (see Section 2.3.1.1). The pitch of the stimulus is related to the pattern of excitation produced by that stimulus (for a pure tone the pitch is generally assumed to respond to the position of maximum excitation) (Moore 1997, p177). An alternative to the place theories, 'temporal' theories, suggests the pitch is related to the temporal patterns of firing within and across neurones. When a neurone is excited, the nerve firings show phase locking, i.e., *"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"* (Moore 1997, p177-178) (see Section 2.3.1.2).

For explaining residue pitch, or virtual pitch, place theories propose that the frequencies of the sinusoidal components of the complex tone are firstly analysis; and then the pitch is derived by pattern recognition from neural signals corresponding to the frequencies of the resolved components (individual partials which are well resolved by the ear; for a harmonic complex sound, they correspond to the lower harmonics) (Moore 1997, p189). Terhardt's (1972; 1974b; 1979) theory of virtual pitch assumed a learning phase for the pattern recognition. He suggested that since we are exposed to speech, which frequently contains harmonic complex tones, from the earliest moments in life, we learn to associate a given frequency component with its subharmonics, which always occur together in harmonic complex tones. After the learning phase is completed, when a harmonic complex tone is presented, the pitch cues corresponding to the subharmonics produced by components coincide at certain values, among which the fundamental frequency at which the largest number of coincidences occurs determines the overall pitch of the sound. Alternatively to Terhardt's theory, Goldstein (1973) supposed a central processor which presumes that all stimuli are periodic and that the spectra comprise

successive harmonics. The processor searches for the harmonic series which provides the 'best fit' to the series of resolved components, in other words, an approximation which allows likely error (Moore 1997, p190-192).

Temporal theories suggest that the pitch of a complex tone is derived from the time intervals between successive nerve firings evoked at points on the BM where are excited by the tone (Moore 1997, p211). The time intervals between firings in a given neurone reflect those between peaks in the temporal structure of the waveform driving that neurone (Moore 1997, p204). In Schouten's (Schouten *et al.* 1962; Schouten 1970) theory, the residue pitches may be perceived resulting from unresolved components (for a harmonic complex sound, these correspond to the upper harmonics which are not well resolved by the ear but interfere on the BM). The value of a residue pitch is determined by the periodicity of the waveform at the point on the BM where the partials interfere, i.e. the total waveform of the unresolved partials (Moore 1997, p194). More strictly, the perceived pitch corresponds to the time intervals between firings occurring at peaks in the fine structure of the waveform (on the BM) close to adjacent envelope maxima (Moore 1997, p195). However, a residue pitch can still be heard when no partials interact. Thus, Moore (1977; 1997) assumed that pitch perception would be based on the temporal pattern of successive impulses evoked by both resolved and unresolved components on the BM. The lower harmonics are revolved, i.e. analysed into effectively separate locations on the BM. The time intervals relate to the frequencies or submultiples of the frequencies of those harmonics. For the higher harmonics, the time intervals between spikes in neurones with higher CFs correspond to the patterns of vibration on the BM interference of a number of harmonics; in other words, neural impulses are derived from each peak in the fine structure of the waveform (Moore 1997, p204-205). Common time intervals are searched for across the different neurones, among which the most prominent ones are selected to determine the pitches. Usually the time interval which is found corresponds to the period of the fundamental component (Moore 1997, p212).

However, neither of the theories can account for all of the experimental results. The pattern recognition models require that one or more partials in a complex sound should be resolvable in order for the perception of a low residue pitch, but a residue pitch may still be heard when none of the individual partials is separately perceptible. The temporal theories could not work for sinusoids at very high frequencies, since phase locking of the stimulating waveform disappears above 4-5 kHz. It is likely that both types of theories are utilized but that their relative importance is different for different frequency ranges and for different types of sounds (Moore 1997, p211). Pitch modelling is further reviewed in Chapter 5.

2.3.5 Timbre

Timbre has been defined by the American Standards Association (ASA 1960) as “that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar”. In a restricted sense, or in a classical view which was first stated by Helmholtz (von Helmholtz 1863) over a century ago, timbre may be considered the subjective counterpart of the spectral composition of sounds (Rasch and Plomp 1982, p13), i.e., the distribution of energy over frequency. A steady-state sound can be described by a multitude of frequencies with particular intensities and relative phases. Later research has shown that temporal characteristics of sounds may have a profound influence on timbre as well, which has led to a broadening of the concept of timbre (Schouten 1968). Both onset effects (rise time, presence of noise or inharmonic partials during onset, unequal rise of partials, characteristic shape of rise curve, etc.) and steady state effects (vibrato, amplitude modulation, gradual swelling, pitch instability, etc.) are important factors in the recognition of an 'auditory object' and hence in the timbre of sounds (Rasch and Plomp 1982, p13).

Unlike loudness or pitch, which may be considered as unidimensional, timbre is a multidimensional attribute of the perception of sounds; i.e., sounds cannot be ordered on a single scale with respect to timbre (Moore 1997, p246). Dimensional research of timbre leads to the ordering of sound stimuli on the dimensions of a timbre space (Rasch and Plomp 1982, p14). The most important factors, or dimensions, of timbre found can be characterized as follows: (a) the shape of spectral energy distribution which includes (1) sharpness, determined by a distribution of spectral energy that has its gravity point in the higher frequency region; (2) compactness, a factor that distinguishes between tonal (compact) and noise (not compact) aspects of sound, in other words, whether the sound is periodic, having a tonal quality, or irregular and having a noise-like character (Schouten 1968; von Bismarck 1974b); (3) the spectral composition of the sounds, i.e. sound spectrum which is multiple dimensions, e.g. the relative level produced by a sound in each critical band (Plomp et al. 1967; Pols et al. 1969); and (b) temporal features which include (4) attack rate of onset; (5) the extent of synchronicity among the various components during onset (Wessel 1979); (6) whether the waveform envelope or any other aspect of the sound (e.g. spectrum or periodicity) is constant, or fluctuates as a function of time and in the latter case what the fluctuations are like (Schouten 1968; Moore 1997, p247).

Among the timbre factors, sharpness, compactness, roughness and fluctuation are reviewed in the next sections.

2.3.6 Sharpness

Sharpness, or brightness, as described above as a dimension of timbre of steady sounds correlated to the spectral energy distribution, is a sensation attribute increased with the upper limiting frequency as well as with slope of the spectral envelope of sounds (von Bismarck 1974a), in other words, caused by high frequency components in a sound (HEAD acoustics GmbH 2011b).

For narrow-band noises, sharpness increases with increasing centre frequency. It increases almost in proportion to centre frequency in critical-band rate at low frequencies below 3 kHz; for higher frequencies, it increases faster than the critical-band rate (von Bismarck 1974a; Zwicker and Fastl 1999, p240). For noises with bandwidth wider than a critical band, sharpness increases if bandwidth increases by increasing upper cut-off frequency, and decreases if bandwidth increases by reducing lower cut-off frequency. In other words, sharpness of a sound increases by adding sound at higher frequencies and decreases by adding sound at lower frequencies (Zwicker and Fastl 1999, p241). For the dependence on sound level, sharpness increases for a level increment from 30 to 90 dB by a factor of two. This effect is small and can be ignored in some cases (Zwicker and Fastl 1999, p239).

Based on overall spectral envelope, a model of sharpness has been developed (von Bismarck 1974a). As discussed above, sharpness is mainly influenced by spectral envelope, but independent of the fine structure of spectrum (Zwicker and Fastl 1999, p239). Spectral envelope can be psychoacoustically represented in the excitation level or the specific loudness versus critical-band rate pattern (Zwicker and Fastl 1999, p241). The increase of sharpness with centre frequency in critical-band rate is taken into account by using a factor, g(z), which is a function of critical-band rate. It is unity for all critical-band rates below 16 Bark but increases for higher critical-band rates from unity to a value of four near 24 Bark, which taking into account that sharpness increases strongly at high frequencies. An equation of sharpness is given following:

$$S = 0.11 \frac{\int_0^{24\,\mathrm{Bark}} N' g(z) z \, dz}{\int_0^{24\,\mathrm{Bark}} N' \, dz} \,\mathrm{acum}\,.$$

In this equation, sharpness (S) is proportionally equal to a weighted first moment (centre of gravity) of the critical-band rate distribution of specific loudness, using the factor, g(z), in which the numerator corresponds to the integral of weighted specific loudnesses (N') over critical-band rate and the denominator is the total loudness (von Bismarck 1974a; Zwicker and Fastl 1999, p242).

Aures (1985) later modified the calculation method of sharpness of von Bismarck, for coordinating with more results of psychoacoustic measurements on sharpness. Firstly, the factor or weight function in the equation of von Bismarck's model has been re-determined. Instead, an exponential function of critical-band rate is used. Furthermore, to correct the sharpness for different loudness, the denominator in the equation is replaced by a weight function of loudness using natural logarithm.

2.3.7 Tonality

Tonality, or compactness, another factor of timbre of steady sounds, indicates whether a sound consists mainly of tonal components or broadband noise (Aures 1985).

Sound with relative large tonality has spectrum of only a line, which corresponds approximately to that of a sinusoidal tone. Broadband noise, such as white noise, has little or no tonality. Narrowband noise components of the spectrum may contribute to a limited extent to tonality if the bandwidth is smaller than a critical band (Terhardt and Stoll 1981; Aures 1985). Thus, bandwidth has an effect on tonality that with increasing bandwidth tonality of band noise decreases. Additionally, tonality is influenced by the frequency for pure tones. It generally decreases with increasing frequency, but increases with increasing frequency at low frequencies below about 500 Hz (Aures 1985). This band-pass characteristic is similar to the phenomenon of spectral dominance in pitch perception (Terhardt *et al.* 1982). Tonality is independent of loudness for loudness in medium or average range.

According to these characteristics, a calculation procedure has been developed (Aures 1985). First, the sinusoidal and narrow-band components of a sound are identified. The tonal sinusoidal components are obtained from a fast Fourier transform (FFT) spectrum of sound according to the pitch analysis method of Terhardt (Terhardt 1979; Terhardt *et al.* 1982). The components are extracted by the detection of local maxima in FFT spectral samples (spectrum is represented by 400 samples with an upper frequency limit of 5 kHz, and thus with sample spacing of 12.5 Hz). They are determined whether the SPL of the spectrum sample is at least 7 dB higher than that of the next three lower and three higher samples. Then a tonal component is represented by group of the seven spectral (Terhardt *et al.* 1982). In addition, the narrow-band noise components of the spectrum are identified in the spectrum from which all sinusoidal components are removed. The areas, which are narrower than a critical band, with relatively high intensities, i.e. the noise levels are at least 7 dB higher than those in adjacent critical bands, are defined as narrow-band noise components (Aures 1985).

In the next step, the sound pressure level excess (SPL excess) or the surplus level is calculated for all the tonal sinusoidal and narrow-band spectral components for evaluation of masking effects according to Terhardt (Terhardt 1979; Terhardt *et al.* 1982). SPL excess is defined as the difference between the SPL of an individual component and that SPL which represents the masking power of the rest of the spectrum (Terhardt 1979). It is the level of the component minus the threshold in quiet, the noise intensity in the respective critical band, and the excitation level produced by the other tonal components (Terhardt *et al.* 1982).

With the bandwidth, frequency (centre frequency) and SPL excess of the tonal components, a weighted function is calculated for each component. Finally, the tonality is calculated by adding all the weighted functions of the components and taking into account the ratio of the loudness of the signals without the tonal components and loudness with the components (Aures 1985; HEAD acoustics GmbH 2011b).

2.3.8 Roughness

When two simple tones of slightly different frequency sounding simultaneous, several perceptual phenomena occur depending on the condition of frequency difference between the two tones. The intensity of the result signal of the two tones is alternately greater and less in regular succession, which result in a periodically fluctuating perceived loudness with a repetition frequency equal to the frequency difference of the two tones (von Helmholtz 1863; Rasch and Plomp 1982, p14-15). The places of maximum intensity are called *beats* (von Helmholtz 1863), if they can be discerned individually by the ear, which occurs if their frequency is less than about 20 Hz. The beats raise a sensation of loudness fluctuation. When the frequency difference is larger than about 20 Hz, the ear is no longer able to follow the rapid amplitude fluctuations individually, and instead a rattle-like sensation called *roughness* occurs (Rasch and Plomp 1982, p15). Beats and roughness only occur if the two simultaneous tones are not resolved by the ear, i.e., the frequency difference is less than the critical bandwidth, due to the interaction, or partial overlap, of their activity patterns on the basilar membrane (BM) (Fishman *et al.* 2000). If the frequency difference is larger than the critical band, the tones are perceived individually with no interference phenomena (Plomp and Levelt 1965; Plomp and Steeneken 1968; Terhardt 1974a).

An alternative approach to roughness is by means of studying amplitude-modulated (AM) stimuli (Terhardt 1974a; Fastl 1977). These two approaches generally produce similar predictions, as proximal frequency components give rise to amplitude beats and

amplitude modulation gives rise to proximal frequency components (i.e. three-component complex) (Terhardt 1974a; Pressnitzer and McAdams 1999). With an AM tone, similarly to simultaneous tones, three different areas of sensation are reached using increasing modulation frequency from low to high. They are the sensation of fluctuation, which reaches a maximum at modulation frequencies near 4 Hz and decreases for higher frequencies, the sensation of roughness, which starts to increase at about 15 Hz and reaches its maximum near modulation frequencies of 70 Hz, and the sensation of hearing three separately audible tones, which increases as roughness decreases (Zwicker and Fastl 1999, p257).

The perception of roughness is confined to modulation frequencies in the region between about 15 to 300 Hz. Roughness strength depends on the frequency and depth of modulation for amplitude modulation (Fastl 1990). The dependency of roughness on the modulation frequency shows a band-pass characteristic, i.e., it reaches a maximum near modulation frequency of 70 Hz - about 50 to 70 Hz depending on the centre frequency, and decreases towards low or high modulation frequencies. With increasing modulation depth, the impression of roughness increases. In addition, it increases with increasing sound pressure level (SPL) to a small extend, by a factor of about 3 for a increase in SPL by 40dB (Zwicker and Fastl 1999). It is interpreted that roughness is determined by the envelope fluctuations of signal within an auditory filter, or a critical band. Most narrow-band noises sound rough even though there is no periodical change in envelope or frequency, because the envelope of the noise changes randomly (Zwicker *et al.* 1979).

Roughness has been therefore modelled based on two main factors that influences, which are frequency resolution and temporal resolution of our hearing system (Zwicker and Fastl 1999, p261). Frequency resolution is modelled by the excitation pattern or by specific loudness versus critical-band rate pattern. Temporal resolution is based on the changes or differences in excitation level (masking depth) or in specific loudness at all places along the critical-band rate scale, i.e. the specific loudness-time function in each channel, taking into account the masking effect produced by strongly temporally varying maskers. An equation of roughness is given as follow,

$$R = 0.3 \frac{f_{\mathrm{mod}}}{\mathrm{kHz}} \int_0^{24\mathrm{Bark}} \frac{\Delta L_{\mathrm{E}}(z)\, dz}{\mathrm{dB/Bark}} \,\mathrm{asper}\,.$$

Roughness (R) is proportional to the product of modulation frequency (f_{mod}) and masking depth (ΔL) summed up across critical bands.

2.3.9 Fluctuation strength

As described in the last section, the hearing sensation of fluctuation will be produced when a tone is modulated at low modulation frequency, up to about 20Hz (Zwicker and Fastl 1999). Similar to roughness, fluctuation strength of amplitude-modulated (AM) pure tone depends on modulation frequency, modulation depth and sound pressure level. The maximum of fluctuation strength occurs at a modulation frequency of around 4 Hz, which corresponds the variation of temporal envelope of fluent speech – 4 syllables/second are usually produced at normal speaking rate. This may be seen as an indication of excellent correlation between speech and hearing system. Fluctuation strength increases approximately linearly with the logarithm of modulation depth. With increasing sound pressure level (SPL), fluctuation strength increases, by a factor of about 3 with a increase of 40dB in SPL (Zwicker and Fastl 1999, p248).

A model of fluctuation strength is proposed based on the temporal variation of the masking pattern or loudness pattern, similar to that of roughness (Zwicker and Fastl 1999, p253-256). Fluctuation strength is calculated from the masking depths of the temporal masking patterns, i.e., the differences between the maxima and the minima in specific loudness, which are integrated along the critical-band rate, and modulation frequency. More details in modelling of fluctuation strength, as well as of roughness strength, are given in Chapter 3 Section 3.3.1.

2.3.10 Auditory temporal processing

Time is an important dimension in hearing, which reflects temporal aspects of the perception of time varying sounds (almost all sounds fluctuate over time) (Moore 1997, p148). This section focuses on temporal resolution, which refers to limits in the ability to detect changes in stimuli over time, i.e. normally not the resolution of changes in the fine structure – the rapid pressure variations in a sound, but in the envelope – the slower overall changes in the amplitude of these fluctuations (Viemeister and Plack 1993).

Since an auditory pattern is a time varying sequence of spectral shapes, changes in time pattern of a sound are generally associated with changes in its magnitude spectrum. Subjects are able to detect the spectral differences, either by monitoring the energy within single critical bands, or by detecting the differences in spectral shape of sound (Moore 1997, p149). Thus, *"in characterizing temporal resolution in the auditory system, it is important to take account of the filtering that takes place in the peripheral auditory system. Temporal resolution depends on two main processes: analysis of the time pattern*

occurring within each frequency channel; and comparison of the time patterns across channels" (Moore 1997, p148).

Regarding within-channel temporal resolution, the threshold for detecting a temporal gap is typically 2-3ms for broadband noises (Moore 1997, p174). Measurements of narrowband sounds generally show that within-channel resolution does not vary markedly with centre frequency (Moore 1997, p175). It is not clear how across-channel temporal resolution is dependent on within-channel processing, but it seems likely that such a dependency exists (Viemeister and Plack 1993). The detailed modelling of temporal resolution is further discussed in Chapter 6.

2.4 Psychology of Music

In von Helmholtz's classic volume "On the Sensations of Tone" (von Helmholtz 1863) published more than a century ago, with the subtitle "As a Physiological Basis for the Theory of Music", Helmholtz indicated the theory of music could be understood as its origin in the perceptual characteristics of our hearing system, although music theory has its own rules apart from the perceptual relevance. With the development of electroacoustic means necessary for psychoacoustical experiments, the relationship between musical-theoretical and perceptual entities has been investigated through much research (Rasch and Plomp 1982).

2.4.1 Consonance and dissonance

In the theory of music, consonant (musical) intervals correspond to simple ratios between the frequencies of the tones, for example, 2:1 (octave), 3:2 (fifth), 5:4 (major third), and 6:5 (minor third). *"When musical notes in these simple ratios are sounded simultaneously, the sound is pleasant, or consonant, whereas departures from simple, integral ratios, tend to result in a less pleasant or even a dissonant sound"* (Moore 1997, p208).

The perceptual (or psychoacoustic) consonance (Plomp and Levelt 1965), distinguished from consonance in a musical situation (musical consonance has its roots in perceptual consonance, but is dependent on the rules of music theory, which, to a certain extent, can operate independent from perception), of an interval consisting of two simple tones depends directly upon the frequency difference between the tones, not upon the frequency ratio (or musical interval). If the frequency separation is very small or large, more than a critical bandwidth (the tones not interfering with each other), the two tones

together sound consonance. Dissonance occurs if the frequency separation is less than a critical bandwidth (Rasch and Plomp 1982, p19). In other words, psychoacoustic consonance is correlated with the absence of roughness, while dissonance is caused by roughness (Plomp and Levelt 1965; Terhardt 1974b). In the situation of two harmonic complex tones sounding simultaneously, roughness can be produced by interference of a great number of harmonics. When the fundamental frequencies are in simple ratios, several of their harmonics coincide, whereas for non-simple ratios the harmonics differ in frequency and produce beating or roughness sensations (Moore 1997, p209). These findings strongly support Helmholtz's (1863) consonance theory (Terhardt 1974b).

Roughness, or interference of harmonics on the BM, may explain at least part of the dissonance; however, it cannot account for the whole of the effect, when considering our preference for certain frequency ratios for both the simultaneous and successive presentation of tones (Moore 1997, p209). Terhardt (1974b) suggested a principle termed tonal meanings in addition to the principle of minimal roughness, which governed the musical consonance. The principle of tonal meanings suggests a learning process, which could also account for the perception of residue pitch as discussed in Section 2.3.4, would account for the perception of musical intervals (Moore 1997, p209). That is, we learn about particular frequency ratios (octave and other musical intervals) by exposure to harmonic complex sounds (usually speech sounds) from the earliest moments in life (Moore 1997, p209).

While there is a psychoacoustic basis for consonance and dissonance judgements (Moore 1997, p210), consonant (or pleasant) sounds are not necessarily preferred to dissonant sounds (Parncutt 1994). Perceived consonance, like most psychological parameters, displays individual differences and depends on cultural experience of listener and musical context presented (Parncutt 1994; Moore 1997, p210).

2.4.2 Rhythm

Rhythm is a general term that refers to the time-dependent properties of events or sound (Brown 1993). Time is not merely a passive medium within which events occur; rather, it acts to shape and determine all phenomena. Thus, rhythm energizes, structures, creates, and expresses temporal quality (Handel 1989, p383). In the psychology of music, rhythm has not been as thoroughly studied as pitch, although rhythmic information is more fundamental to music cognition than pitch information. It is probably because in addition to peripheral processes in the nervous system like those concerned with pitch, more complex central processes are involved for temporal and rhythmic information (Dowling

and Harwood 1986, p178-179). A rhythmic perception is a subjective experience that relies on the context of the phenomenal experience of rhythm, that is, no component of acoustic signal can uniquely specify the rhythm (Handel 1989, p384).

As contrasted with *rhythm* in general, a number of specific terms are used in describing temporal properties of musical events. *Duration* refers to "the psychological correlate of time". *Beat* refers to "a perceived pulse marking off equal durational units". *Tempo* refers to "the rate at which beats occur" (Dowling and Harwood 1986; Brown 1993). *Meter* refers to the alternation of subjectively strong and weak beats (Handel 1989, p391).

Rhythmic organization is an inherent part of all human activity; moreover, it exists at all levels of activity (Handel 1989, p383). This means that rhythms emerge from diverse rhythmic levels (Handel 1989, p383) and that each rhythmic level is in turn dependent on every other level (Handel 1989, p390-391). In both music and speech, each level becomes a pattern of beats (Handel 1989, p392), and meter is perceived as layered and thus is described as hierarchical (Martin 1972; Handel 1989, p391). The process of creating the hierarchy determines the relationships among the levels and among the groups at each level (Handel 1989, p391).

The underlying attributes of rhythmic experience have been studied with different experimental procedures (Gabrielsson 1973). Three groups of attributes or dimensions appear to be common and important: the structure dimension which distinguishes rhythm on the basis of meter, degree of accent on the first beat, type of underlying pattern, clearness of marked basic patterns, uniformity-variation (simplicity-complexity); the movement dimension which distinguishes rhythm on the basis of rapidity/tempo, forward movement/motion, movement/motion; and the emotion dimension (Handel 1989, p456-458).

2.4.3 Music and emotion

The relationship between emotions and music is as complex as emotional phenomena in general (Dowling and Harwood 1986, p202). *"The easiest emotional responses to understand are those tied directly to biological survival, for example, the reaction of fear in the face of a life-threatening situation"* (Dowling and Harwood 1986, p202). Thus, while reactions to music are not obviously of such direct biological significance, it is possible that emotional responses may be somehow relatively direct to environmental sound.

This section reviews, on the one hand, organisation of emotions and, on the other hand, contribution of factors in musical structure to the emotional reactions / perceived emotional expression.

2.4.3.1 Emotions

Psychologists describe affect either by categories or low dimensional spaces. Each representation is supported by a large body of psychology research. Categorical approaches cluster emotional descriptors into a set of independent factors or dimensions, such as displeasure, distress, depression, excitement, etc., varying between six and twelve depending on research (Hevner 1936; Borgatta 1961; Lorr *et al.* 1967).

However, there was evidence that rather than being independent, these affective dimensions are interrelated in a highly systematic fashion. In the work by Russell and Thayer (Russell 1980; Thayer 1989), two-dimensional circumplex models were proposed, to place all emotional descriptors in a valence-arousal (V-A) space, where the amount of arousal (activation-deactivation) and valence (pleasure-displeasure) is measured along the vertical and horizontal axis, respectively. In the spatial model, affective states can be represented a circle in the following order: pleasure (0°), excitement (45°), arousal (90°), distress (135°), displeasure (180°), depression (225°), sleepiness (270°), and relaxation (315°) (Russell 1980). Some studies have expanded this approach to develop three-dimensional spatial models, although the semantic nature of the third dimension is subject to speculation and disagreement (Bigand *et al.* 2005).

2.4.3.2 Music factors and emotions

There is a strong association between music and emotions. Music can express or represent emotions as well as arouse or induce them (Gabrielsson 2001). Musical emotional expression has been given considerable attention in empirical research since about one hundred years ago, while emotional reactions to music have been less studied (Gabrielsson 2001; Gabrielsson and Lindström 2001). However, *"it may be that often the emotion represented is also the emotion induced, through this is not always the case"* (Dowling and Harwood 1986, p203).

This section therefore focuses on studies of the relationship between different factors in musical structure and perceived emotional expression, using a variety of methods. These music factors include tempo, loudness, pitch, mode, melody, rhythm, harmony, and various formal properties.

Tempo is usually considered the most important among factors affecting emotional expression in music (Gundlach 1935; Hevner 1937; Rigg 1964; Gagnon and Peretz 2003).

The term tempo usually refers to perceived pulse rate, but may not always have the same meaning. Perceived speed may also be influenced by note density, the number of notes per unit of time (e.g. per second) (Gabrielsson and Lindström 2001). Studies indicate that fast tempo / high note density may be associated with various expressions of activity/excitement (Hevner 1937; Watson 1942), happiness/joy/pleasantness (Hevner 1937; Rigg 1940; Watson 1942; Wedin 1972; Krumhansl 1997; Peretz *et al.* 1998; Balkwill and Thompson 1999), potency, surprise, flippancy/whimsicality, anger, uneasiness (Gundlach 1935), and fear. Slow tempo / low note density may be associated with various expressions of calmness/serenity (Hevner 1937), peace (Balkwill and Thompson 1999), sadness (Hevner 1937; Watson 1942; Wedin 1972; Krumhansl 1997; Peretz *et al.* 1998; Balkwill and Thompson 1999), dignity/solemnity (Gundlach 1935; Hevner 1937; Rigg 1940; Wedin 1972), tenderness, longing (Rigg 1940), boredom, and disgust. While both fast and slow tempo may thus be associated with many different expressions, the perceived expression in each case is highly dependent on the context; that is, presence and level of other structural factors (Gabrielsson and Lindström 2001). In terms of the valence-arousal model (as reviewed in Section 2.4.3.1), fast tempo/ high note density is generally associated with high activation, slow tempo/ low note density with low activation, while both of them may be associated with either positive or negative valence (Gabrielsson and Lindström 2001).

For mode, major mode may be associated with expressions as happiness/joy (Hevner 1936; Wedin 1972; Crowder 1985; Krumhansl 1997; Peretz *et al.* 1998), graceful (Hevner 1936), serene (Hevner 1936), and solemn; minor mode with expressions as sadness (Hevner 1936; Wedin 1972; Crowder 1985; Krumhansl 1997; Peretz *et al.* 1998), dreamy, dignified (Hevner 1936), tension, disgust, and anger. Perceived expression depends on the context. Moreover, major mode is not a necessary condition for perceived happiness; a piece in minor mode in fast tempo may very well sound happy (Gabrielsson and Lindström 2001). While differences between fast and slow tempo are mainly associated with difference in activation in the valence-arousal model, differences between major and minor mode are mainly associated with difference in valence, positive or negative (Gabrielsson and Lindström 2001).

In terms of loudness (or intensity), loud music may be associated with various expressions of intensity/power (Wedin 1972), excitement (Watson 1942), tension (Krumhansl 1996), anger, and joy; soft music with softness (Wedin 1972), peace (Watson 1942), tenderness, sadness, solemnity, and fear. On the whole, loud music seems to be associated with high activation, soft music with low activation (Gabrielsson and Lindström 2001). Large variations of loudness/intensity may suggest fear, small variations happiness or activity. Rapid changes in loudness may be associated with

playfulness, pleading (Watson 1942), and fear (Krumhansl 1997), few or no changes with sadness, peace, and dignity.

For pitch, high pitch may be associated with expressions such as happy, graceful, serene, dreamy (Hevner 1937), and exciting (Watson 1942), and also with surprise, potency, anger, fear, and activity. Low pitch may suggest sadness, dignity/solemnity, vigour, and excitement (Hevner 1937; Watson 1942; Wedin 1972), as well as boredom and pleasantness, – such an apparent contradiction may depend on the musical context. Large pitch variation may be associated with happiness, pleasantness, activity, or surprise, small pitch variation with disgust, anger, fear, or boredom. Wide melodic (pitch) range may be associated with joy (Balkwill and Thompson 1999), whimsicality, uneasiness (Gundlach 1935), and fear (Krumhansl 1997), and narrow range with sad (Balkwill and Thompson 1999), dignified, sentimental, tranquil, delicate, and triumphant (Gundlach 1935).

For other factors, simple and consonant harmony may be associated with expressions such as happy/gay (Hevner 1936; Watson 1942), relaxed, graceful, serene, dreamy (Hevner 1936), dignified (Hevner 1936; Watson 1942), serious, and majestic (Watson 1942); complex and dissonant harmony with excitement (Hevner 1936; Watson 1942), tension (Krumhansl 1996), vigour (Hevner 1936), anger, sadness (Hevner 1936; Watson 1942), and unpleasantness (Wedin 1972). Sharp amplitude envelope with rapid attack and decay may be associated with anger, happiness, surprise, and activity, and round envelope with tenderness, sadness, fear, disgust, boredom, and potency (Gabrielsson and Lindström 2001). High complexity of musical form (melodic/harmonic/rhythmic) may be associated with tension (Krumhansl 1996), sadness (Balkwill and Thompson 1999), melancholy, depression, anxiety, or aggressiveness, and low complexity with relaxation, joy, or peace (Balkwill and Thompson 1999).

In sum, perceived emotional expression in music is rarely or never exclusively determined by a single factor, but is always a function of many factors. The influence of a certain factor may depend on how it is combined with other factors, that is, on the interaction between factors (Gabrielsson and Lindström 2001). Tempo and loudness, in particular among the factors, seem to show most distinct effects, – increase in either of them results in higher activation, and decrease results in lower activation. Also generally the activation dimension seems more salient and easier to judge than the valence dimension (Schubert 2004; Leman *et al.* 2005; Gomez and Danuser 2007).

2.5 1/f Noise in Music and Soundscape

All physical measurements are ultimately limited by fluctuations or "noise" in either the system being measured or the measuring apparatus (Voss 1979). The common noises found in nature fall into three general classes according to their spectral densities, which varies as $1/f^{\gamma}$, where f is the frequency and $0 \leq \gamma \leq 2$. For white noise, γ is equal to zero; whereas γ is equal to 2 when the parameter does a random walk, called brown noise; when γ is close to 1, the type of variability is called $1/f$, or "pink" noise.

Vacuum tubes, carbon resistors, semiconducting devices, continuous or discontinuous metal films, ionic solutions, films at the superconducting transition, Josephson junctions, nerve membranes, sunspot activity, and the flood levels of the river Nile all exhibit what is known as "1/f noise" (Voss and Clarke 1978).

2.5.1 Spectral density and time correlations

The spectral density (also known as power spectrum), as a characterization of average behaviour of a quantity varying in time, is a measure of the mean square variation in a unit bandwidth centred on the frequency f (Voss and Clarke 1978; Voss 1979). An alternative characterization of the average behaviour is autocorrelation function, which is a measure of how the fluctuating quantities at times t and $t + \tau$ are related. Spectral density and autocorrelation function are related by the Wiener-Khintchine relations (Reif 1965).

In the case that fluctuating quantity is characterised by a single correlation time, from the Wiener-Khintchine relations, it is possible to show that spectral density is "white" (independent of frequency) in the frequency range corresponding to times over which the fluctuating quantity is independent; and is a rapidly decreasing function of frequency, usually $1/f^2$, in the frequency range over which the quantity is correlated (Voss and Clarke 1978). A quantity with a 1/f spectral density cannot, therefore, be characterized by a single correlation time. In fact, the 1/f spectral density implies some correlation in the fluctuating quantity over all times corresponding to the frequency range for which spectral density is 1/f-like. In other words, a quantity with a white spectral density (white noise) is uncorrelated with its past, showing rapid uncorrelated changes, and has the most random appearance; a quantity with a $1/f^2$ spectral density ($1/f^2$ noise) depends very strongly on its past, showing only slow changes, and is the most correlated; and a quantity with a 1/f spectral density (1/f noise) has an intermediate behaviour, with some correlation over all times, yet not depending too strongly on its past, and exhibits a

balance between randomness and correlation on all time scales (Voss and Clarke 1978; Voss 1979).

2.5.2 1/f noise in music and speech

Voss and Clarke (1978) studied the spectral densities of audio signal, audio power (varying closely with loudness), and frequency fluctuations of single records of music and radio stations (of durations of approximately 12 h, greater than a single record). While the spectral density of audio signal from music is far from 1/f, the spectral density of the slowly varying quantities, i.e., fluctuations in the audio power and frequency, of many musical selections varies approximately as 1/f. For a classical station, the spectral densities of audio power and rate of zero crossings of audio signal (frequency fluctuations) exhibit a smooth 1/f dependence; for a rock station and a jazz and blues station, the spectral densities are 1/f-like down to frequencies corresponding to the average selection length, and flatten at lower frequencies. Among several music pieces by different composers, although the measured audio power and frequency fluctuations for all the pieces show an increasing spectral density at lower frequencies, individual differences can be observed (Voss and Clarke 1978).

For English speech, the spectral density of audio power for a news and talk station is 1/f-like. However, the spectral density of zero crossing rate has a quite different behaviour, *"characterized by two correlation times: The average length of an individual speech sound, roughly 0.1 s, and the average length of time for which a given announcer talks, about 100 s"* (Voss and Clarke 1978).

De Coensel et al. (2003) repeated the analyses of amplitude and pitch fluctuations in music and speech, using 4 classical pieces, with the duration that varies between half an hour and one hour, and a speech fragment of radio. For pressure amplitude and pitch fluctuations of the music fragments, the resemblance to 1/f behaviour is obvious. Speech fragments behave slightly different. In the region 0.1 to 1Hz the spectrum of pitch is almost flat; at lower frequencies, the 1/f dependence is recovered (De Coensel *et al.* 2003).

Voss and Clarke (1978) explained the 1/f-like spectral density of quantities associated with music and speech as the result of a critical balance between predictability and novelty (De Coensel *et al.* 2003). It was later interpreted as music being an imitation of the temporal fluctuation of self-organized criticality (SOC) system that seems so common in nature (De Coensel *et al.* 2003). SOC is generally believed as a source of 1/f behaviour of complex system (Bak *et al.* 1987).

2.5.3 1/f noise in soundscapes

De Coensel et al. (2003) studied 1/f noise in outdoor soundscapes, using sound fragments of 15 minutes each recorded monaurally in rural and urban soundscapes, in terms of dynamics of both loudness and pitch variations. They found that the expected 1/f behaviour, previously found in music, also appeared in many soundscapes, although with certain deviations.

2.5.3.1 Rural soundscapes

Based on recordings of 6 different rural locations selected as silent areas in Flanders, Belgium (L_{Aeq} between 40 and 45 dBA) and at different times of day, 1/f behaviour of typical rural soundscapes were studied (De Coensel *et al.* 2003). The sound events in these recordings included distant traffic, airplanes, farm noises, farm animals, and birds. The A-weighted sound level, loudness, and pitch power spectra "in general show 1/f behaviour, but deviate much more from this characteristic than music". It is common to find 1/f in the frequency interval [0.2Hz, 5Hz], which corresponds to a time interval between 200 ms and 5 s and therefore associates to characteristic of the sound fluctuations within the source itself. In the frequency interval [0.002Hz, 0.2Hz], which corresponds to the time interval between 5 s and about 10 min and is therefore influenced mainly by fluctuations between sources, *"all rural soundscapes have more slow variations in loudness and pitch than expected in the case of SOC"*, which indicates the predictability. Between A-weighted sound level and loudness, *"loudness spectra seem to show a clearer trend than A-weighted pressure spectra"*. It may result from low frequency sound *"that is probably caused by distant man-made noises such as traffic"*, for which *"A-weighting does not adequately remove these unheard components as Zwicker loudness does"* (De Coensel *et al.* 2003).

2.5.3.2 Urban soundscapes

Urban soundscapes were recorded in the city of Ghent, Belgium, including 12 recordings in residential area, open square, shopping street, tourist attracting embankment, park area, and blocks of flats in open green setting. The general 1/f trend in A-weighted sound pressure, loudness, and pitch is obvious. Similar to the rural soundscapes, in the frequency interval [0.2Hz, 5Hz], the power spectra of loudness and pitch follow closely to the typical 1/f frequency dependence. On longer time scales, i.e., in the frequency interval [0.002Hz, 0.2Hz], some of the urban soundscapes show the same characteristic (more slow variations in loudness and pitch than SOC) as rural soundscapes, although they may be much louder on the average. In other urban

soundscapes, 1/f or even a flatter spectrum is observed, that is, self-organization may be more common here (De Coensel *et al.* 2003).

2.5.4 Relation of 1/f noise to people's perception and evaluation, and descriptors for the temporal structure of soundscape

By presenting stochastic compositions, in which the frequency and duration of each note were by determined by white, 1/f and $1/f^2$ noise, to several hundred people, Voss and Clarke (1978) found that the 1/f music was judged by most listeners to sound pleasing. Those generated by white noise sources (music with a flatter spectrum) sounded too random, chaotic, and unpredictable, while those generated $1/f^2$ noise (with a steeper slope) sounded too correlated, predictable, and hence boring and dull (Voss and Clarke 1978; De Coensel *et al.* 2003).

By extension of this finding, Botteldooren et al. (2006) proposed a descriptor for the temporal structure of urban soundscape, which "measures the similarity of its spectrum of loudness (and pitch) fluctuations to those typical for music". Based on both the measured average slope of the spectrum and the quadratic deviation from the best-fitted straight line, an indicator for degree of music-likeness was constructed from a fuzzy set membership function. Since the amplitude and pitch spectrum of music has an approximate 1/f or a 1/f-like behaviour, the fuzzy set membership function was constructed on the basis of the probability distribution of slopes and deviations that are found in music, by analysing the spectra of 15 samples of music of different genres. This descriptor proposed was then compared to L_{A5} - L_{A95} as a classical indicator of dynamics. Correlation between both descriptors, based on 31 soundscapes, indicated that the new descriptor proposed probed a different dimension of the soundscape (Botteldooren *et al.* 2006).

2.6 Summary

In this chapter, in addition to the area of soundscape that attempts to join together researches in the various areas, three independent areas of sonic studies are reviewed, i.e., psychoacoustics, psychology of music, and 1/f noise dynamic.

From the review of soundscape, it is clear that although soundscape evaluations are influenced by the factors of sound, environment, and people, which cover the scope of soundscape, almost all the relevant soundscape research show a similar tendency of human listeners of preferring natural sounds, rejecting mechanical sounds, and having

neutral attitudes towards human sounds. It indicates a need to explore the factors for such sound preferences, that is, the link between sound sources and objective measures of sound.

The review of the scientific areas of psychoacoustics, psychology of music, and 1/f noise dynamic, suggests the possibility of applying the objective parameters in the three areas to measurement of environmental sound in soundscapes. These parameters include loudness, pitch, timbre, rhythm, and 1/f noise, which have been applied or suggested in the fields of music, industrial acoustics design, and some soundscape studies, as they reflect subjective sensations evoked by sound, concerning with the way that sound is perceived and/or cognised in the auditory system.

Moreover, as music and soundscape are closely related, it is expected that the knowledge of music psychology, more specifically, the effects of music on emotion, etc., would benefit study of relationships between soundscape and emotion.

Chapter 3
Methodology

In order to explore the differences between nature and urban sound and among different categories of sound, the characteristics of different categories of sound are analysed in terms of three aspects: psychoacoustic parameters that have been recommended in previous soundscape research, additional psychoacoustic parameters that have mainly been applied in music perception, and dynamic indicator that has been used for analysing music and soundscapes.

In this chapter, the methodology for this research is described. An illustration of the methods can be seen in Figure 3.0.1. First, the classification of sound, or say categories, and relevant definitions used in this research are discussed in Section 3.1. Then, Section 3.2 describes the collection of sound samples. Section 3.3 describes the methods for calculating each aspect of the parameters. Finally, the statistic methods for analysing the results and methods for automatically identifying the sound categories are discussed in Section 3.4.

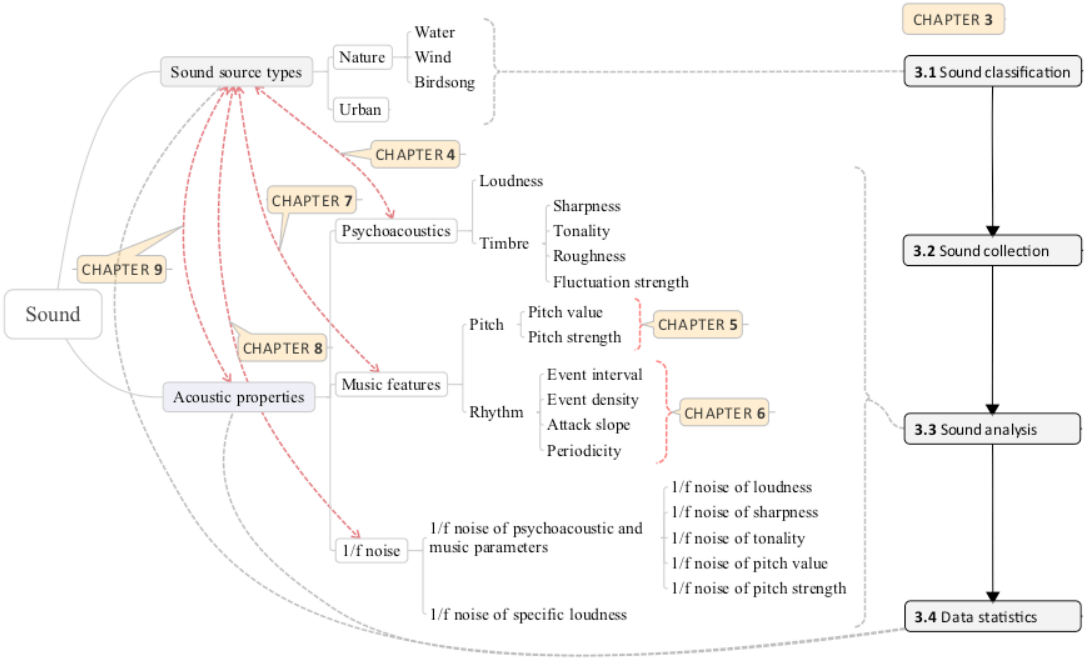


Figure 3.0.1 Method illustration

3.1 Sound Classification

3.1.1 Sound classification in literature

Sounds may be classified in many ways, e.g. according to their physical characteristics (acoustics) or the way in which they are perceived (psychoacoustics); according to their function and meaning (semiotics and semantics); or according to their emotional or affective qualities (aesthetics) (Schafer 1977, p133; Payne *et al.* 2009).

For generic classifications, Schafer (1977) categorised the main themes of a soundscape by distinguishing between keynote sounds, signals, and soundmarks. Keynote is a musical term; it is the note that identifies the key or tonality of a particular composition. Keynote sounds may not always be heard consciously, the fact that they are ubiquitously there suggests the possibility of a deep and pervasive influence on human's behaviour and moods. Signals are foreground sounds and they are listened to consciously. Schafer confined some of the signals must be listened to because they constitute acoustic warning devices: bell, whistles, horns and sirens. The term soundmark is derived from landmark and refers to a community sound that is unique or possesses qualities which makes it specially regarded or noticed by the people in that community. Dubois et al. (2006) derived two broad categories from linguistic analysis of subjects' free descriptions of soundscapes: source events, which can be attributed to an identified source, and background noise, where no specific events can be discriminated.

Using the free-sorting tasks, Dubois et al. (2006) found sounds were often categorised together as they were produced by similar sources, or by similar movements or actions. Everyday sounds have been mostly categorised by researchers into 'human', 'nature' and 'mechanical' (Payne *et al.* 2007). Schafer (1977) developed classification that had been used for one of the sub-projects of the World Soundscape Project, an extended card catalogue of descriptions of sound from literary, anthropological and historical documents. Sounds were classified into six main categories: natural sounds, human sounds, sounds and society, mechanical sounds, quiet and silence, and sounds as indicators. Each main category contains some subcategories, e.g., natural sounds have the subcategories of sounds of creation, apocalypse, water, air, earth, fire, birds, animals, inserts, fish and sea creatures, and sounds of seasons; while sounds and society has the subcategories of rural soundscapes, town soundscapes, city soundscapes, maritime soundscapes, domestics soundscapes, sounds of trades, professions and livelihoods, sounds of factories and offices, sounds of entertainments, music, ceremonies and festivals, parks and gardens, and religious festivals. Brown et al. (2011) put a common framework for classification of all sound sources in any acoustic environment forward

standardization in soundscape assessment. The taxonomy of the acoustic environment has been constructed in terms of categories of places—indoor, outdoor—and within the outdoor environment: urban, rural, wilderness and underwater.

3.1.2 Soundscape classification framework

As discussed in Chapter 2, a soundscape framework can be divided into three aspects, i.e., sound, environment, and human; thus, soundscapes can be classified according to any of these aspects. In order to obtain an integrated classification system of soundscape, in this section, a multi-dimensional classification framework is proposed, where a particular soundscape is classified according to all of these aspects simultaneously. That is, the multi-dimensional classification framework is composed of each soundscape aspect as an independent dimension for classification. For example, the soundscape of birdsongs in a park perceived by people enjoying the environment is classified by birdsongs for sound, park for environment, and enjoying for human. Another example can be the soundscape of an urban square perceived by people passing by, which is classified with undefined sound – which might be mixed sound sources that may be found in an urban square, such as fountain, voice and traffic –, square for environment, and passing by for human.

Here, it gives a simple illustration of this framework, while a more complete or complex framework can include the sub-aspects in each of the aspects, or be further extended by additional aspects, with the same principle. The sub-aspects, as discussed in Chapter 2, are force causing the object's vibration and vibrating object for the aspect of sound, medium through which sound propagates and environment effecting the sound's propagating for the aspect of environment, and perception of human's ear and auditory system and cognition of higher level system in brain for the aspect of human; that is, the six steps in the whole process from sound being produced to being perceived by human. Changes in any of the sub-aspects, including force, object, medium, environment effect, perception, and cognition, would change the whole process, and thus the classification.

In the context of this framework, the sounds being studied in this research are classified according to different sound types in the aspect of sound, and for any environment and any state or background of people.

3.1.3 Sound definition and classification

While this research concerns the sound aspect of soundscape, which intends to link sound types with subjective sensations of human's auditory system in the process, the sound

samples used in this study, consider the most frequently heard sound sources in soundscapes of everyday life in urban area, including sounds from both nature and human activity/facility (Brown *et al.* 2011). As reviewed in Chapter 2 Section 2.2.7, natural sounds are generally preferred by human, while mechanical and human sounds are either rejected or have low levels of preference. However, people's attitudes towards human related sounds differ greatly between people and between different types of sound (Yang and Kang 2003; Yang and Kang 2005b). Thus, in this study, two general sound categories are considered, which are sounds that are natural and all other sounds that are non-natural in everyday soundscapes.

As sound can be seen as being produced by force that causes the object's vibration and vibrating object as discussed above, in this research, natural sound is defined as the sound that caused by nature, including natural phenomenon, wildlife, etc., and comes from the natural materials or objects; in other words, both the force causing the object's vibration and vibrating object are natural. Based on this definition, sounds in this research are classified into natural sounds and all sounds that besides natural sounds, of which either or both the force and object is not natural, and is termed as urban sounds in this study.

Table 3.1.1 Sound category and subcategory, and number of recordings and 30-seconds segments contained in each category and subcategory

Category		Number of Recordings	Subcategory	Number of Recordings	Number of 30s Segments
Natural Sounds	Water Sounds	34	Stream	6	35
			Small river	4	66
			Medium river	4	44
			Wave on shingle	13	158
			Wave on sand	4	33
			Wave into cove	3	31
	Wind Sounds	23	Wind in deciduous trees	4	47
			Wind in coniferous trees	10	117
			Wind in heath	9	119
	Birdsongs	28	Birdsong in woodland	10	162
			Birdsong in heathland and grassland	6	62
			Birdsong in moorland and wetland	4	99
			Birdsong in farmland and village	6	73
			Birdsong in coastal	2	33
Urban Sounds		17	Church bell	2	5
			Fountain	2	5
			Street music	2	15
			Machine	4	9
			Traffic	3	37
			Human voice	3	5
			Footsteps	1	4
Total		102		102	1159

As shown in Table 3.1.1, the natural sounds consist of three categories: water sounds, wind sounds, and birdsongs. Each category is further classified into several subcategories. Water sounds are further classified into six subcategories: stream, small river, medium river, wave on shingle, wave on sand, and wave into cove. Wind sounds are classified into three subcategories: wind in deciduous trees, wind in coniferous trees, and wind in heath. Birdsongs sound are classified into five subcategories: birdsongs in woodland, in heathland and grassland, in moorland and wetland, in farmland and village, and in coastal. The human related sounds, or urban sounds as termed in this paper, contain the subcategories of church bells, fountains, street music, street machines, traffic, human voice and footsteps.

3.2 Sound Recording Collection

A wide range of sound recordings of single sources is used as sound samples in the research, since the current study lays emphasis on the differences among various sound sources. However, environmental sound is usually mixed with multiple sound sources, which brings the difficulty that it may take large amount of time to record single source sounds. Thus, in order to obtain a large number of samples, the recordings used in this study are either recorded by the authors or collected from multiple databases.

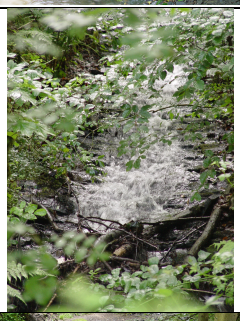
As part of the recordings is collected from various sources, the recording equipment and filtering settings vary. The absolute sound pressure level (SPL) is available for the recordings from some databases and the recordings made by the authors, but not for other recordings. Verifications are consequently made to examine the effects on the psychoacoustic analysis for sounds.

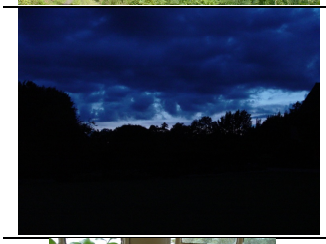
3.2.1 Sound sample recordings collection

3.2.1.1 Sound recording

The recordings made by the author were recorded in England countryside, natural parks, and urban area in the summer of 2009, which include sounds of stream, wind, rain, swan, and traffic. Information of part of the recordings, including recording place and date, are summarised in Table 3.2.1.

Table 3.2.1 Recording information of the sound recordings

	File name	Sound type	Recording place	Recording date	Filter setting
	Porter Brook_087	Stream	Porter Brook, Sheffield, South Yorkshire	6 Jul 2009	-
	Porter Brook_089	Stream	Porter Brook, Sheffield, South Yorkshire	6 Jul 2009	-
	Porter Brook_090	Stream	Porter Brook, Sheffield, South Yorkshire	6 Jul 2009	-
	Porter Brook_092	Stream	Porter Brook, Sheffield, South Yorkshire	6 Jul 2009	-
	Wind & water_094	Wind	Cold Hiendley Reservoir and Wintersett Reservoir, Wakefield, West Yorkshire	23 Jul 2009	-
	Wind_095	Wind	Cold Hiendley Reservoir and Wintersett Reservoir, Wakefield, West Yorkshire	23 Jul 2009	-
	Wind on grass_096	Wind	Cold Hiendley Reservoir and Wintersett Reservoir, Wakefield, West Yorkshire	23 Jul 2009	-

	File name	Sound type	Recording place	Recording date	Filter setting
	Wind_097	Wind	Cold Hiendley Reservoir and Wintersett Reservoir, Wakefield, West Yorkshire	23 Jul 2009	100Hz 12dB/Oct
	Wind_099	Wind	Haw Park Wood, Wakefield, West Yorkshire	23 Jul 2009	100Hz 12dB/Oct
	Traffic_104	Traffic	Weston Manor, Weston-on-the-Green, Oxfordshire	6 Aug 2009	-
	Rain_106	Rain	Weston Manor, Weston-on-the-Green, Oxfordshire	7 Aug 2009	-
	Rain & traffic_107	Rain	Weston Manor, Weston-on-the-Green, Oxfordshire	7 Aug 2009	-
	Rain & traffic_108	Rain	Weston Manor, Weston-on-the-Green, Oxfordshire	7 Aug 2009	-
	Swan_109	Bird	Blenheim Park, Woodstock, Oxfordshire	8 Aug 2009	-

The stream sounds were recorded at Porter Brook, a river in the City of Sheffield in South Yorkshire. It flows eastward from its source inside the Peak District National Park (Wikipedia 2013). The recordings were made along the river from Hunter's Bar to Forge Dam. The wind sounds were recorded in the area of Haw Park Wood, Cold Hiendley Reservoir and Wintersett Reservoir (Wakefield Council 2013). Haw Park Wood was designated a Local Nature Reserve, located 4 miles south east of Wakefield, West Yorkshire. The woodland is now dominated by conifers with areas of broadleaved trees. Some traffic noise and the rain sounds were recorded around Weston Manor (2013), which is a historic building dating back to the 11th Century situated in Weston-on-the-Green on the edge of the Cotswolds countryside, near Oxford, Oxfordshire. The fountain sounds and swan sounds were recorded at Blenheim Park, a monumental stately home and a World Heritage Site situated in Woodstock, Oxfordshire, surrounded by over 2,000 acres of landscaped parkland, the great lake, and formal gardens (Blenheim Palace 2013). More traffic noises were recorded near the Netherthorpe Road in Sheffield.

The recordings were made in the mono mode, using the equipment of Behringer B-2 PRO condenser microphone and Fostex FR-2 field memory recorder. Behringer B-2 PRO condenser microphone boasts a frequency response from 20 Hz to 20 kHz with a slight presence boost, about 5dB, in the range around 12 kHz and a small dip, about 2dB, at around 5 kHz (when the omni-directional pickup pattern is used) (Behringer 2013). Among the three pickup patterns: omni, cardioid (for picking up the source signal while rejecting off-axis sound), and figure eight, the omni pickup pattern was selected for capturing sound in all directions. All the recordings were made without attenuator or low frequency cut filter in the microphone. With Fostex FR-2 field memory recorder, the sound signals were recorded in digital files in Broadcast WAV file format (BWF) (Fostex 2013), with 16 or 24-bit quantisation and 48kHz sample rate. Some of the wind recordings were made with a high-pass filter in the recorder, with the cut-off frequency at 100 Hz and filter slope at 12 dB/oct. Part of this detail information of the recordings is displayed in Table 3.2.1.

The sound pressure level (SPL) was monitored during recording with an A-weighted sound pressure level meter. At the beginning of each recording, a loud pulse of sound was produced by the author and recorded. Then, the recordings were processed on computer by adjusting the whole recording’s volume (level) to equal the instant A-weighted SPL of the pulse moment of the recording to the maximum value measured by the level meter over the duration. In this way, the levels of the recordings are equal to those of the actual sounds in field.

As sound in environment is usually composed of multiply sound sources which are mixed together, however, it brings the difficulty to record sounds with single sound

source for a relatively long duration. Of most of the recordings that made by the authors, the durations without any other type of sound intrudes are less than 30 seconds. Thus, only a few of the recordings are used for the analysis in the study. To solve this problem, the recordings used in this study were further widely collected from databases.

3.2.1.2 Sound recording collections from databases

In order to obtain a large number of samples, the recordings used in this study were mainly collected from multiple databases, including the British Library Sound Archive (2013a), published CDs (Cusack 2001), and the Positive Soundscape Project database (Davies *et al.* 2007), and were made from 1994 to 2010.

The natural sound recordings, including water sounds, wind sounds and birdsongs, were mainly collected from British Library Sound Archive in the British wildlife recordings section. The natural sound recordings were made for the most part in nature reserves and wild location around Britain (British Library Sound Archive 2013b). Various mixed combinations of recording equipment were used for these recordings, which include Sennheiser MKH20/30/40/418S microphones, with Canford/ Filmtech/ Electro Acoustique Appliquée (EAA) PSP2 preamplifiers, and Aiwa HHB DAT/ Tascam DAP1 DAT/ Sony TCD D7 DAT/ Nagra Ares-BB+/ Sound Devices 702 recorders. The cut-off frequency settings for low frequency filtering vary from 40/80/100/200Hz 6/12dB/Oct or none, according to different sound types. All the recordings are stored in uncompressed wave digital format with 44.1kHz, 16bit or 96.0kHz, 24bit quantisation. The absolute sound pressure level (SPL) was not available for the recordings; as a result, the SPL range for each subcategory of sound was measured in similar sound environments to be compared with, as discussed in the following section.

The urban sound recordings were mainly collected from the published CD "Your Favourite London Sounds" and the Positive Soundscape Project database. The CD includes human activity/facility sound recordings made in the urban areas of London, while Positive Soundscape Project database includes those made in the urban areas of Manchester. Recording equipment and filter setting information was not available for all or part of these recordings. The absolute SPL was available for the recordings from the Positive Soundscape Project database, but not for the recordings from the CD.

Since the recordings differ in their durations, each recording is divided into a number of 30-second segments for the analyses. Research reported that recognition process of human subjects on a couple of soundscapes took average time of 20 seconds (Peltonen *et al.* 2001). Also, numbers of studies on music emotion and information retrieval used segments of 30 seconds for analyses (Kim *et al.* 2010). All the recordings used in the study, including both the recordings made by the authors and collected from databases,

were stored in uncompressed wave digital format, and converted into the same sample rate and size of 44.1kHz 16bit in this study, both of which are equal or lower than the original rate and size of the recordings. The pilot study of this research used the part of these recordings in the compressed digital file format of MP3. The performances of these two audio formats, i.e. Wave and MP3, for psychoacoustic analyses are compared in Section 3.2.4.

One hundred and two recordings, of about 700 minutes in total, have been used in this study. In Table 3.1.1 the number of recordings and the number of segments contained in each subcategory are shown. It is noted that whereas the natural sound recordings have only single sound source be heard, for the urban sound recordings, although a specific sound is dominant, some general background sounds may still be heard. It is expected it results from that different sound sources are usually mixed together in urban areas rather than in natural/rural environments, which also brings a difficulty for recording urban sounds of single source. Consequently, the recordings that are available to collect are less for urban sounds than natural sounds, which results that the number of recordings in the urban sound category is less than that in the natural sound category, and that there are more detailed subcategories in the natural category than in urban. However, when considering four categories, i.e., water sound, wind sounds, birdsongs, and urban sounds, the numbers of recordings are relatively balanced across the categories. As each category has covered most frequently heard sounds in everyday life in that category, the unequal numbers of recordings across categories may not significantly affect the final results.

3.2.2 Sound pressure level measurement of each category of sound

The absolute sound pressure level (SPL) was available for the recordings from the Positive Soundscape Project database and the recordings made by the authors, but not for other recordings. To estimate the SPL range for various typical sounds as shown in Table 3.1.1, a series of 30-seconds L_{Zeq} of sound environments was measured in England with a 01dB digital sound and vibration level meter, where Z- or Zero frequency weighting is defined in IEC 61672 (2003) as a linear frequency response of 10Hz to 20kHz. Also, for a number of the categories or subcategories, the SPL ranges were accessed by the SPLs calculated from the recordings by the authors. The measured/calculated SPL range, displayed in Table 3.2.3, was compared with those of the collected recordings of each subcategory, and only those sound samples with a SPL difference of less than 5dB were retained for analyses, ensuring that the analysed samples were in reasonable sound level ranges. In the table, the absolute SPLs were available for the recordings in certain

subcategories, thus no additional measurements were made. Table 3.2.3 also shows the place and date of the measurements or recording, as well as the average SPL of recordings retained in each subcategory.

3.2.3Influence of low-frequency cut filtering of sound on psychoacoustic analysis

As the recordings were collected from various sources, the recording equipment and filtering settings vary. The cut-off frequency settings for low frequency filtering vary from 40/80/100/200Hz 6/12dB/Oct or none, according to different sound types. In this section, the influence of low frequency filtering on psychoacoustic analysis is studied using four typical sounds, including water, wind, bird and traffic. These 30-seconds sound samples were first recorded without filtering and then filtered by 100Hz 12dB/Oct low-cut in the software of ArtemiS (see Section 3.3.1). The recordings before and after filtering were analysed in terms of the psychoacoustic parameters, i.e. loudness, sharpness, tonality, roughness, and fluctuation strength (also see Section 3.3.1).

Table 3.2.2 shows the average results over the duration under both conditions and also the differences between the two. It can be seen that the differences between the two conditions in terms of these parameters are generally relatively small compared to their absolute values, except for fluctuation strength of wind sounds and tonality of traffic noise. These results suggest that the low-cut filtering generally does not significantly influence the majority of results of psychoacoustic analysis in this study. Thus, no attempt has been made to compensate for these filtering effects.

Table 3.2.2 Average values of the psychoacoustic parameters before and after low-cut filter of four types of sound

Type	File name		Loudness (sone)	Sharpness (acum)	Tonality (tu)	Roughness (asper)	Fluctuation Strength (vacil)
Water	Porter Brook_090_4 (0.00-30.00 s)	Before	15.5	2.97	0.019	2.44	0.0101
		After	15.4	2.99	0.024	2.43	0.0100
		Difference	-0.1	0.02	0.005	-0.01	-0.0001
Wind	Wind_095 (54.00-84.00 s)	Before	8.9	2.10	0.023	1.43	0.0156
		After	8.7	2.13	0.029	1.44	0.0083
		Difference	-0.1	0.03	0.006	0.01	-0.0073
Bird	022A-W1CDR0001344-0700P0 (15.00-45.00 s)	Before	18.5	4.41	0.064	2.09	0.0505
		After	18.4	4.43	0.069	2.08	0.0505
		Difference	-0.1	0.02	0.005	-0.01	0.0000
Traffic	Traffic_128 (700.00-730.00 s)	Before	18.8	1.63	0.074	2.26	0.0104
		After	17.8	1.68	0.117	2.20	0.0115
		Difference	-1.0	0.05	0.043	-0.06	0.0011

Table 3.2.3 Measured/calculated SPL of category of sound

Category	Sound type	File name	Recording/measurement place	Recording/measurement date	SPL (dB)	L_{Zeq} (dB)	Subcategory	SPL AVE (dB)
Water Sounds	Stream and river	Stream_076	Grindleford, Derbyshire	24/06/2009	55.3	-	Stream	64.1
		Stream_077		24/06/2009	61.4	-	Small river	74.7
		Stream_078		24/06/2009	53.6	-	Medium river	71.8
		Stream_079		24/06/2009	64.3	-		
		Stream_080		24/06/2009	70.7	-		
		Stream_082		24/06/2009	58.5	-		
		Porter Brook_087_1 (0.00-20.00 s)	Porter Brook, Sheffield, South Yorkshire	06/07/2009	75.4	-		
		Porter Brook_087_2 (0.00-20.00 s)		06/07/2009	76.1	-		
		Porter Brook_089_1 (0.00-20.00 s)		06/07/2009	75.8	-		
		Porter Brook_089_2 (0.00-30.00 s)		06/07/2009	76.3	-		
		Porter Brook_090_1 (0.00-50.00 s)		06/07/2009	73.8	-		
		Porter Brook_090_3 (0.00-20.00 s)		06/07/2009	74.0	-		
		Porter Brook_090_4 (0.00-30.00 s)		06/07/2009	72.8	-		
		Porter Brook_092_1 (0.00-40.00 s)		06/07/2009	71.8	-		
	Sea wave	-	Beach, Liverpool, Merseyside	24/06/2011	-	69.7	Wave on shingle	59.1
		-		24/06/2011	-	74.1	Wave on sand	63.3
		-		24/06/2011	-	63.0	Wave into cove	70.6
		-		24/06/2011	-	70.8		
Wind Sounds	Wind	Wind_095	Cold Hiendley Reservoir and Wintersett Reservoir, Wakefield, West Yorkshire	23/07/2009	82.8	-	Wind in deciduous trees	64.3
		Wind_097		23/07/2009	59.5	-	Wind in coniferous trees	72.3
		Wind_099	Haw Park Wood, Wakefield, West Yorkshire	23/07/2009	63.2	-	Wind in heath	66.8
Birdsongs	Bird	-	Peak district, Derbyshire	10/04/2011	-	56.7	Birdsong in woodland	62.5
		-	Western Park, Sheffield, South Yorkshire	17/04/2011	-	68.3	Birdsong in heathland and grassland	59.9
							Birdsong in moorland and wetland	55.9
							Birdsong in farmland and village	56.7
							Birdsong in coastal	64.5
Urban Sounds	Church bell	-	-	-	-	-	Church bell	67.8
	Fountain	-	Peace garden, Sheffield, South Yorkshire	17/04/2011	-	75.1	Fountain	70.8
		-	Barkers Pool, Sheffield, South Yorkshire	17/04/2011	-	73.5		
	Street music	-	-	-	-	-	Street music	81.2
	Machine	-	-	-	-	-	Machine	79.4
	Traffic	-	-	-	-	-	Traffic	69.8
	Voice	-	Lydgate Park, Sheffield, South Yorkshire	18/04/2011	-	70.0	Human voice	71.4
		-		18/04/2011	-	70.1		
	Footsteps	-	The Grange, Sheffield, South Yorkshire	28/06/2011	-	57.8	Footsteps	59.2
		-		29/06/2011	-	58.6		

3.2.4 Influence of sound formats of recordings on acoustic and psychoacoustic analysis

The pilot study in this research used the free access to the recordings in British Library Sound Archive in the digital file format of MPEG Layer-3, commonly referred to as MP3. Thus, the applicability of MP3 format sound recordings for psychoacoustic analysis is studied before the analysis, which would be also useful for potential soundscape research afterwards, since MP3 is a widely used digital audio compression format.

Moving Pictures Expert Group (MPEG) was formed in 1988 to establish standards for the coded representation of moving pictures and associated audio stored on digital storage media (ISO/IEC 13818-3 1998). MPEG-1/2 (ISO/IEC 11172-3 1993; ISO/IEC 13818-3 1998) standardises the generic coding system and consists of three operating modes called layers, with increasing complexity and performance from Layer-1 to Layer-3. MPEG Layer-3 (MP3), with the highest complexity, is optimised to provide the highest quality at low bit-rates (around 128 kbit/s for a stereo signal) (Brandenburg 1999). It works by reducing the accuracy of certain parts of sound that are considered to be beyond the auditory resolution ability of most people, on the basis of psychoacoustic models of perceptual limitation of human hearing system, or more precisely, auditory masking in the critical bands. The lossy compression greatly reduces the amount of data required to represent the audio recording and still sounds indistinguishable from the original uncompressed signal.

In this section, the differences between recordings in the uncompressed audio format of Wave and compressed format of MP3 for psychoacoustic analysis are analysed, using a traffic noise and a rain noise. Both of the sound samples, of a duration of 60 seconds, were recorded by the author and stored in the uncompressed Wave format, with sample rate and size of 44,100 Hz 24 Bits, and mono mode. These sound samples were then converted from Wave format into MP3 format, with sample rate and size of 44,100 Hz 16 Bits and mono mode, using the software of Sony Sound Forge 8.0 (2013), with three transform qualities, i.e., fastest encode, medium, and highest quality, all of which are examined here. The MP3 format sound samples have a bit rate of 128kbps, instead of 1058kbps for the Wave format samples. The spectrum, sound pressure level (SPL), and the psychoacoustic parameters, including loudness, sharpness, tonality, roughness, and fluctuation strength, of each sample in both formats are calculated, using the software of ArtemiS (see Section 3.3.1).

3.2.4.1 The difference between Wave and MP3 format sound recordings in spectrum analysis

Figure 3.2.1 shows the difference between spectra (analysed using FFT method) of the traffic noise recording in Wave and MP3 formats, including the three transform quality levels for MP3 (fastest encode, medium, and highest quality). From the figures, it can be seen that, comparing to Wave format, for the low quality MP3 format, the levels don't change much from frequency of 10 Hz to around 15,000 Hz, and beyond 15,000 Hz the difference increases to 10 to 20dB. For the medium and high quality MP3 formats, the differences of spectra are in a similar tendency; both the differences are very small from 10 Hz to around 18,000 Hz, and are about 35 to 40 dB beyond 18,000 Hz. In Figure 3.2.1 (d), it focuses on the difference between the Wave and medium quality MP3 formats in the frequency range of 10 Hz to 18,000 Hz. It can be seen that the level difference is very small through the range, within about ±0.5dB. The comparison between Wave and MP3 formats in spectrum analysis is further made using a rain noise recording. Here, the medium quality of MP3 format is used, the results of which are shown in Figure 3.2.2. It shows that the result is very similar to that of the traffic noise. That is, between the two formats, the difference is very small, within about ±0.5dB, in the range of 10 Hz to around 18,000 Hz, and is about 35 to 40 dB beyond 18,000 Hz. These results suggest that spectrum analysis using MP3 format sound is reliable in the frequency range of 10 to 15,000 Hz for low quality level, and is reliable in the range of 10 to 18,000 Hz for medium/high quality level of MP3 format.

3.2.4.2 The difference between Wave and MP3 format sound recordings in psychoacoustic analysis

The differences between recordings of Wave and MP3 formats in analysing SPL and psychoacoustic parameters are examined, using the traffic and the rain sound samples. The results are shown in Table 3.2.4, in terms of the average values for each of the parameters of the two formats, as well as the average, maximum, and minimum values of the differences.

For the traffic noise, three transform quality levels of MP3 format sound are examined. The results in Table 3.2.4 show that the difference between the average values of the two formats is small compared to the absolute value for the acoustic and psychoacoustic parameters and thus acceptable, including all the three quality levels. Among the three quality levels, the difference of the highest quality is smallest, while that of the fastest encode is relatively bigger. For the maximum and minimum of the differences varying with time, taking medium quality MP3 format for example, it can be seen from Table 3.2.4 that the difference between the values of level is from about 0.03 to

-0.49, indicating that the error range of level is in about ±0.5. Similarly, the approximate error ranges of loudness, sharpness, tonality, roughness, and fluctuation strength are ± 0.4, ±0.03, ±0.1, ±0.04, and ±0.0005, respectively. Comparing to the absolute value, for loudness, the error range is relatively very small and would not effect much. For sharpness and roughness, the error range may just be acceptable. For fluctuation strength, the range is relatively a bit large, and thus result calculated from MP3 format sound may not be accurate. For tonality, the range is large and even larger than the average value; as a result, MP3 format sound recordings would not be used for calculating the tonality of sound.

Table 3.2.4 Average values, maximum, and minimum of the differences between Wave and MP3 formats in terms of SPL and psychoacoustic parameters

				Level dB [SPL]	Loudness (phon)	Sharpness (acum)	Tonality (tu)	Roughness (asper)	Fluctuation Strength (vacil)
Traffic_104 (184.00-320.00 s) (0.00-60.00 s)	Wav	Average		63.01	70.98	1.868	0.02340	1.494	0.002629
	MP3 Low (fastest encode)	Average		62.89	70.80	1.857	0.02213	1.486	0.002674
		Difference	AVE	-0.11	-0.19	-0.011	-0.00127	-0.008	0.000045
			MAX	0.15	0.17	0.019	0.08392	0.017	0.000429
			MIN	-1.43	-0.56	-0.043	-0.12296	-0.057	-0.000398
	MP3 Medium	Average		62.95	70.84	1.857	0.02295	1.486	0.002677
		Difference	AVE	-0.06	-0.14	-0.011	-0.00045	-0.008	0.000048
			MAX	0.03	0.01	0.005	0.09840	0.018	0.000484
			MIN	-0.49	-0.34	-0.025	-0.08849	-0.037	-0.000295
	MP3 High (highest quality)	Average		62.95	70.85	1.857	0.02281	1.487	0.002656
		Difference	AVE	-0.05	-0.13	-0.011	-0.00059	-0.007	0.000027
			MAX	0.03	0.01	0.006	0.07595	0.014	0.000634
			MIN	-0.48	-0.30	-0.027	-0.08072	-0.030	-0.000204
Rain_106 (128.00-196.00 s) (0.00-60.00 s)	Wav	Average		62.71	71.49	2.190	0.02904	1.517	0.004889
	MP3 Medium	Average		62.65	71.35	2.178	0.02890	1.507	0.004897
		Difference	AVE	-0.06	-0.15	-0.012	-0.00014	-0.010	0.000008
			MAX	0.00	-0.03	0.003	0.08164	0.014	0.000266
			MIN	-0.17	-0.27	-0.030	-0.07369	-0.038	-0.000229

Table 3.2.4 also shows the differences between the Wave and MP3 formats of the rain noise recording, where the MP3 format sample was transformed using the medium quality. The values of both the formats varying with time and the differences between them for all the parameters are shown in Figure 3.2.3. It can be seen that, similar to the results of the traffic noise above, the differences between the average values are small, while the error ranges are relatively small for level, loudness, sharpness, and roughness, but larger for fluctuation strength and tonality. These results suggest that MP3 format sound recordings generally can be used for acoustic and psychoacoustic analysis in terms of average values, but for values varying with time, it may be acceptable for calculating level, loudness, sharpness and roughness, but not for tonality and fluctuation strength.

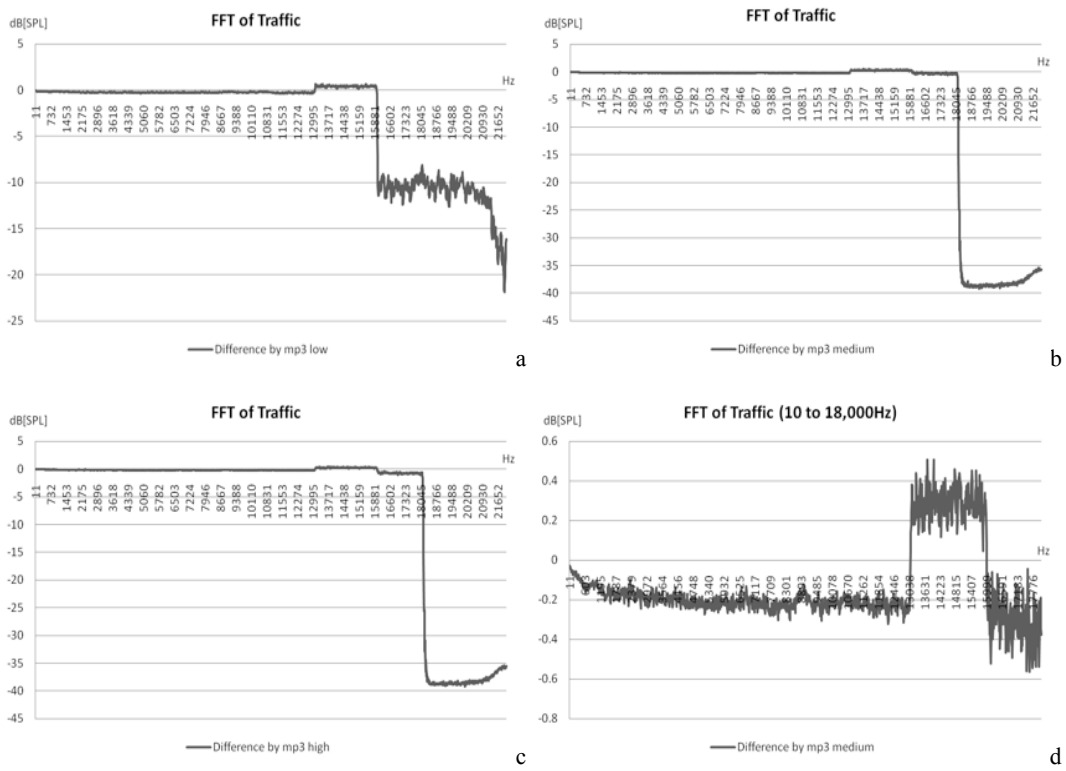


Figure 3.2.1 Differences between the results of FFT analyses of a traffic sound recording in the Wave and MP3 formats, (a) low quality level MP3 format, (b) medium quality level MP3 format, (c) high quality level MP3 format, and (d) medium quality level MP3 format in the frequency range of 10 to 18,000 Hz

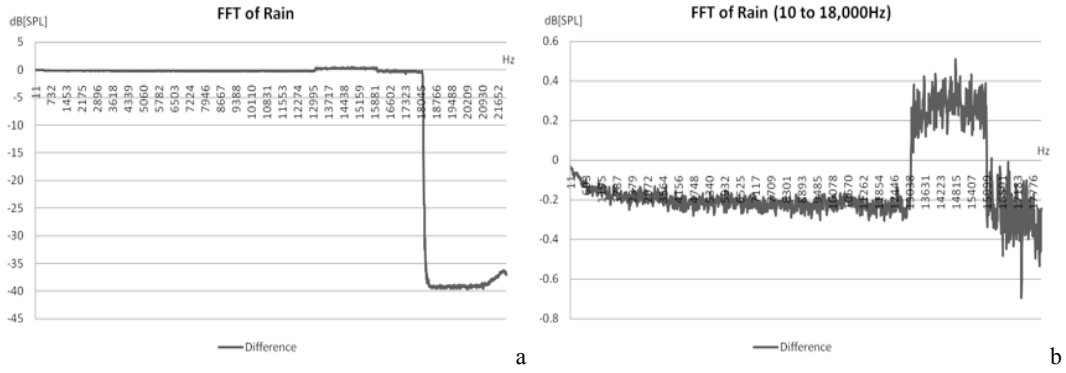


Figure 3.2.2 Difference between the results of FFT analyses of a rain sound recording in the Wave and MP3 formats, (a) medium quality level MP3 format in the full frequency range, and (b) medium quality level MP3 format in the frequency range of 10 to 18,000 Hz

It is noted that the MP3 format sound samples used here may be somehow different if transformed by different software. The comparisons here give a general result on the

reliable frequency range in spectrum analysis and error ranges in calculation of the psychoacoustic parameters.

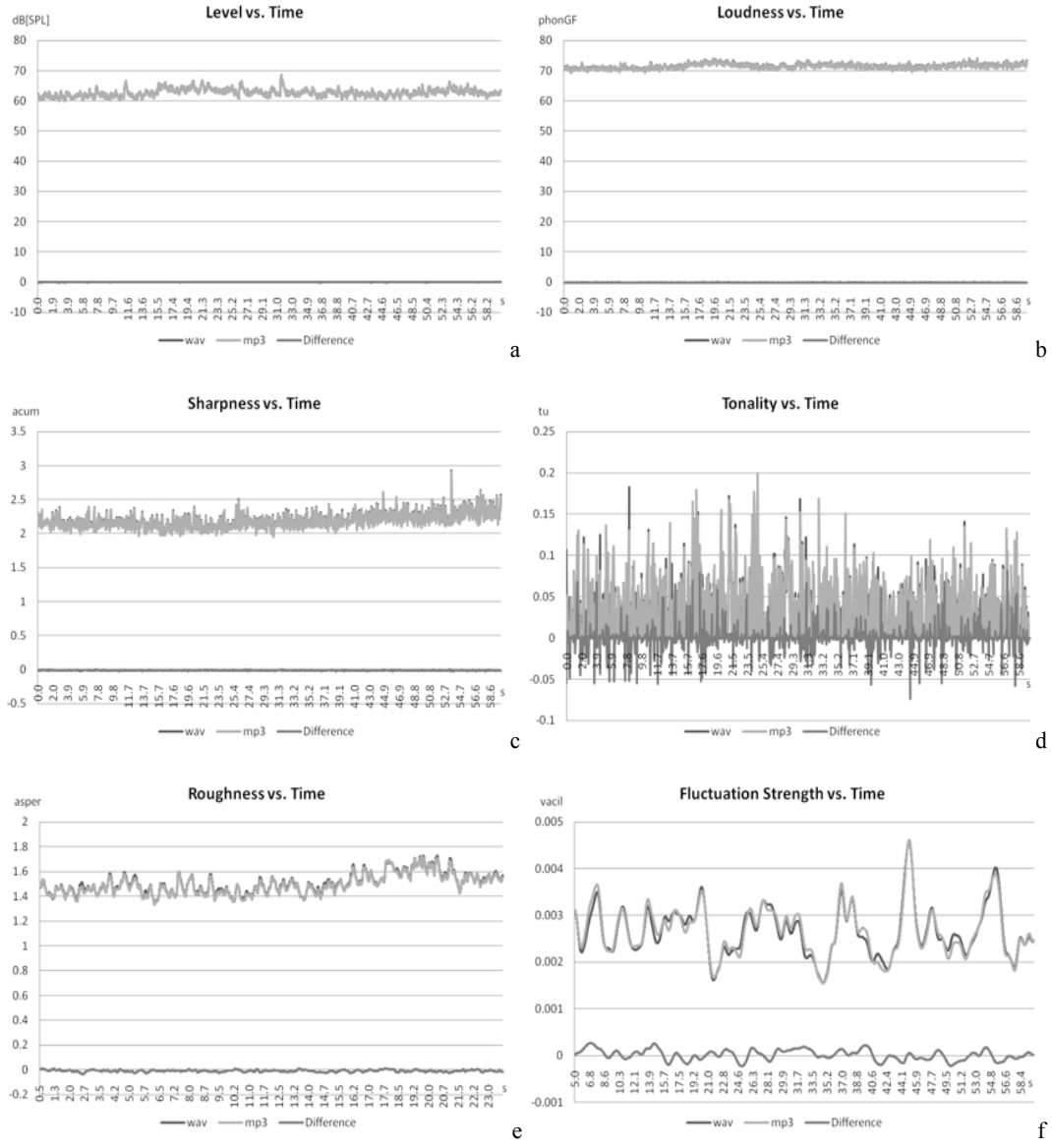


Figure 3.2.3 Psychoacoustic parameters' values varying with time of a rain sound recording in both the Wave and MP3 formats and the differences between them, (a) level, (b) loudness, (c) sharpness, (d) tonality, (e) roughness, and (f) fluctuation strength

3.3 Sound Analysis

The sound recordings are analysed in this research in terms of three aspects: psychoacoustic parameters that have been recommended in previous soundscape research, including loudness, sharpness, tonality, roughness, and fluctuation strength;

additional psychoacoustic parameters that have mainly been applied in music perception; and dynamic indicator of 1/f noise that has been used for analysing music and soundscapes.

In the section, the methods for calculating the psychoacoustic parameters including loudness, sharpness, tonality, roughness, and fluctuation strength are first described. Then it describes the software that used for the analysis of additional psychoacoustic parameters, or say music features for distinguishing with the previous ones. Finally, the method for calculating 1/f noise behaviour is described. The detailed methods for calculating music features are discussed in Chapters 5 and 6.

3.3.1 Psychoacoustic analysis

The psychoacoustic parameters, including loudness, sharpness, tonality, roughness, and fluctuation strength, together with sound pressure level (SPL) are analysed for the recordings over time. The calculations of the six parameters are made using ArtemiS 10 (Advanced Research Technology for Measurement and Investigation of Sound and Vibration) (HEAD acoustics GmbH 2011a).

The calculation of the psychoacoustic parameters is made for each 30-seconds segment of the recordings. For each segment, the six parameters are calculated in terms of the average (AVE), standard deviation (STD), maximum (MAX) and minimum (MIN) over the time of 30 seconds. The results of these indices for each recording are then calculated by averaging all the segments of the recording, considering the influence of unequal durations across the recordings. For example, recordings with longer durations may have larger chance to show higher maximum values. The average of maxima of segments of a recording presents the estimated maximum value in a segment time.

It is noted that there may be different algorithms from these used in this study, for the calculation of roughness, as well as other parameters, although currently there is only a limited number of relevant standards. However, given the main purpose of this study is to distinguish various sounds rather than determine the absolute values of those parameters, it is expected that the influences of calculation methods would not significantly change the main conclusions of this study.

3.3.1.1 Loudness

For psychoacoustic analysis, the calculation of loudness is according to *FFT/ISO 532B* method in ArtemiS (HEAD acoustics GmbH 2011b), among four different methods available including *DIN*, *FFT/ISO 532B*, *Filter/ISO 532B*, and *FFT/HEAD*. The method

used here is based on ISO standard 532B (1975) of method for calculating loudness, which standardised a graphic procedure according to Zwicker (Zwicker *et al.* 1984). The procedure established a specific loudness pattern from third-octave levels of signal, and calculated loudness by summing the specific loudness (also see Chapter 2 Section 2.2). This procedure was also described and specified in the German standard DIN 45631 (1991) with a computer program, which approximates the graphic procedure. The calculation of loudness in ArtemiS is according to this program. For psychoacoustic analysis of loudness in Chapter 4, the third-octave levels are calculated by an FFT of signal, rather than through a filter bank (*Filter/ISO 532B* method). To estimate loudness over time, FFT window length of 4096 samples and a Hanning window of 50% overlap are used. Free sound field loudness diagram is chosen, rather than diffuse field, for analysing environmental sounds in this study.

For 1/f noise analysis of specific loudness (see Section 3.3.3), the calculation of specific loudness is according to *DIN* method, based on German standard DIN 45631. While ISO 532 A/B standardized only the loudness calculation for stationary noise, DIN 45631/A1 (2010) also described the calculation of time-dependent loudness, estimating the temporal effects of loudness by means of filters. (The *DIN* calculation method is identical to the *Filter/ISO 532B* method, except that *DIN* automatically uses 6^{th} order filters.) Since DIN 45631/A1 standard was not released when the author did the psychoacoustic analysis, the calculation of loudness over time for psychoacoustic analysis is based on ISO 532B, while for 1/f noise analysis of specific loudness, the calculation is based on the new standard DIN 45631/A1.

3.3.1.2 Sharpness

Similarly to loudness, a number of algorithms are available for calculating sharpness, which include *DIN 45692*, *Aures*, and *von Bismarck* methods. As reviewed in Chapter 2 Section 2.3.6, the procedure developed by Von Bismarck (1974a) calculates sharpness based on the distribution of specific loudness throughout the critical band rate, while it refers to sounds of equal loudness, that is, the influence of the absolute loudness upon sharpness is not taken into consideration. DIN 45692 (2009) standardised a calculation method similar to that developed by von Bismarck. Aures (1985) corrected von Bismarck's method that, in addition, the influence of loudness is taken into account. Here, sharpness is calculated based on the algorithm according to Aures. As sharpness calculation is based upon the specific loudness, the same loudness algorithm for psychoacoustic analysis as above is used.

3.3.1.3 Tonality

The calculation of tonality is according to the method of Aures (1985), in which it followed Terhardt's procedure for extraction of tonal components (Terhardt *et al.* 1982; Aures 1985). As reviewed in Chapter 2 Section 2.3.7, the tonality calculation takes into account the bandwidth, centre frequency and SPL excess of the tonal components, and the ratio of the loudnesses of signal without and with the tonal components (Aures 1985; HEAD acoustics GmbH 2011b). Here in ArtemiS, the short-term spectrum for the calculation of tonality is obtained with an FFT analysis with window length over 4096 sampling points and a Hanning window of 50% overlap.

3.3.1.4 Roughness

The roughness calculation uses the method of "Roughness vs. Time" analysis function in ArtemiS software, which *"calculates partial roughness from the modulation depths of partial signal bands and adds them up to determine the total roughness"* (HEAD acoustics GmbH 2011c). The signal is first subdivided into 24 partial bands by a linear-phase filter bank; a partial roughness is calculated from the envelope of each partial band signal, which is obtained by a Hilbert transformation. The envelope is then filtered with infinite impulse response (IIR) filters, modelling the dependency of roughness on the modulation frequency. The modulation depth is calculated by the ratio of the power of the constant component (with integration time of 100ms) of partial band signal after the IIR filtering and the power before the filtering. The partial roughness is proportional to the modulation depth; it also takes into account factors representing the dependencies of roughness on frequency position of the partial band, and on sound pressure level. The total roughness is equal to the sum of all partial roughness (HEAD acoustics GmbH 2011c).

It is stated, *"one problem with this method of roughness calculation is that the analysis of signals with un-modulated noise yields roughness values that are much too high compared to the actual perception"* (HEAD acoustics GmbH 2011c). Although ArtemiS provided an additional roughness algorithm based on the hearing model according to Sottek (Sottek 1994; Sottek *et al.* 1994) in order to avoid the problem, only the former method is available in the current package. However, this study mainly aims to examine the differences between various sounds in the parameters rather than the absolute value. Since all the sound samples used in this study are recorded in everyday environment, without modulation, it is expected the relative difference between sound types would not be significantly influenced.

3.3.1.5 Fluctuation strength

The algorithm for fluctuation strength calculation is similar to the hearing model of roughness of Sottek (Sottek 1994; Sottek *et al.* 1994) (i.e., "Hearing Model Roughness vs. Time" analysis function in ArtemiS, but it *"has been adapted in a way that the maximum of the fluctuation strength is obtained at 4Hz instead of 70Hz as for the roughness"* (HEAD acoustics GmbH 2011c)). First, before *"the signal is subdivided by a filter bank with parallel, overlapping band-pass filters"*, a filtering takes place to account for the influence of the outer and middle ear. Here, 24 band-pass filters are used, namely, 1/1 Bark resolution is chosen for the distance between the centre frequencies of adjacent filters. Similarly to the roughness calculation method, envelopes of the partial band signals are obtained, using the Hilbert transformation. For the next steps, the envelopes are reduced to take the threshold in quiet into account, low-pass filtered, distorted using an exponential function with an exponent of 0.125, and calculated the autocorrelation function. Then, the partial fluctuation strengths are determined by a high-pass filtering and a weighting. The combination of the low-pass and high-pass filters models the typical band-pass characteristic regarding the relationship between fluctuation strength and modulation frequency. Both the cut-off frequencies of the low-pass and high-pass filters, and the weighting are frequency-dependent, accounting for the influence of frequency position of the analysed partial band on fluctuation impression. The total fluctuation strength is calculated by integrating the partial fluctuation strengths (HEAD acoustics GmbH 2011c).

3.3.2 Music features analysis

Additional psychoacoustic parameters that have previously mainly been applied in music perception are examined in their applicability in soundscape research, and then analysed for the recordings of environmental sound. To be distinguished from the previous psychoacoustic parameters discussed in Section 3.3.1, these additional psychoacoustic features are termed as music features in this study.

A few music information retrieval (MIR) software packages are available to analyse the music features of sound, which include Marsyas (Tzanetakis and Cook 2000) and MIRtoolbox (Lartillot and Toiviainen 2007). MIR, which is an interdisciplinary research area, has grown rapidly in the last ten years. It extracts music features from audio signals to classify music recordings based on their semantic content, for example genre, instrument, and emotion, to deal with the challenge of searching, retrieving, and organizing huge numbers of recordings.

Marsyas (Music Analysis, Retrieval and Synthesis for Audio Signals) is an open source software framework for audio analysis, synthesis, and retrieval, and has been widely used for MIR applications (Percival and Tzanetakis 2009). The flexible framework contains a variety of existing building blocks (written in C++) of published algorithms in audio signal processing and pattern recognition, and can be extended with new building blocks (Tzanetakis and Cook 2000). MIRToolbox is a Matlab toolbox dedicated to the extraction of musically related features from audio recordings, which includes around 50 audio and music features extractors and statistical descriptors. It has been developed within the context of a European Project called "Tuning the brain for music" with interdisciplinary collaboration, which is related to the investigation of the relation between musical features and music-induced emotion (Lartillot and Toiviainen 2007).

As all these algorithms in the software packages are originally developed for music/speech analysis and information retrieval, their applicability and performance for soundscape study is first examined and compared in a pilot study, using 11 environmental sounds with single sound source and 2 soundscape sounds with mixed sound sources, representing typical natural and urban sounds. The 11 environmental sound recordings are sounds of stream, river, sea waves, wind, birdsong, fountain, church bells, street music, street machines, traffic, and voice. The 2 soundscape recordings are soundscape on a street with clock and traffic, and that in a park with fountain and geese (also see Chapter 5 Section 5.1). These 13 recordings were made with a mono channel, a duration of 30 seconds, and a sample rate of 44,100 Hz (16 bit), which are compatible with both the software packages. Comparing these two packages, since MIRToolbox provides more music features for analysis with more algorithm options, which are generally desirable for the current intention of looking for the applicable algorithms of music features to soundscape research, MIRToolbox is used for analysing music features of recordings following in this study (Yang and Kang 2011).

The music features and corresponding algorithms applicable to soundscape research are further studied and discussed detailedly in Chapters 5 and 6, as well as statistic indices of the features for analysing environmental sounds in this study.

3.3.3 1/f noise analysis

As reviewed in Chapter 2 Section 2.5, in order to examine the 1/f behaviours of fluctuations of instantaneous loudness and pitch in music and speech, Voss and Clarke (1978) studied the spectral densities of audio power and "instantaneous" frequency of

music records and radio stations. To measure instantaneous audio power, the audio signal was squared, and filtered with a 20-Hz low-pass filter. Instantaneous frequency was measured by the rate of zero crossings of the audio signal, which was also smoothed by a 20-Hz low-pass filter before the spectral density was measured. Spectral densities of the fluctuating quantities were measured using a fast Fourier transform (FFT) algorithm that simulated a bank of filters. A log-log plot of the spectral density of the audio power fluctuations, and of the zero crossing rate, showed the 1/f behaviour.

In the research of 1/f noise in soundscape, De Coensel et al. (2003) studied the spectral densities of fluctuations of A-weighted level, loudness (based on Zwicker's model) and instantaneous pitch, using 15-minutes sound fragments recorded in rural and urban soundscapes. "The instantaneous pitch was approximated by counting the number of zero transitions in 10 ms intervals" similarly to the method used by Voss and Clarke (Voss and Clarke 1978; De Coensel *et al.* 2003). The curves obtained in the log (amplitude) versus log (frequency) domain of the spectral densities were locally averaged over a symmetric interval (De Coensel *et al.* 2003). Since for both music and soundscapes a critical point could be identified on the curve of spectral density around a few seconds, the time interval of interest was split between time structure at the micro-scale, which is typically associated to variations within one acoustic event, and macro-scale, time structure at which is caused by the succession of acoustic events (Botteldooren *et al.* 2006). The 1/f noise descriptors included average slope of spectrum and its deviation from a straight line in time both intervals.

In this study, spectral densities of fluctuations of the parameters of psychoacoustic and music features discussed in the last two sections (Sections 3.3.1 and 3.3.2) are first analysed. These parameters include loudness, sharpness, tonality, and pitch. As 1/f noise measures dynamic of sound in terms of a given parameter, the parameters that reflect the variation of sound with time are not included, e.g. roughness and fluctuation strength. Whereas in both the studies of Voss and Clarke (1978) and De Coensel et al. (2003), instantaneous pitch was measured by the rate of zero crossings of the audio signal, in this study, pitch is calculated according to the model of temporal theories of pitch perception of the auditory system as discussed later in Chapter 5. Here, 1/f noise can be seen as a statistic index of the psychoacoustic and music parameters, in addition to the statistic indices that have been used for analysis as in the last two sections, such as mean and stand deviation.

Second, 1/f noise behaviour of fluctuation of audio power in each critical band is examined, in order to approximately simulate the fluctuations along different places of the basilar membrane in cochlea. The fluctuation of amplitude of signal can be computed by root-mean-square (RMS) of the amplitude, amplitude envelope, or specific loudness in

each critical band. These methods are compared with 13 sound samples that are the same as those used in Chapters 5 and 6 (see Chapter 5 Section 5.1.3). The RMS and envelope in each critical band are calculated in Matlab program with MIRtoolbox as described in Chapter 6, while the specific loudness is analysed using ArtemiS software. As discussed in Section 3.3.1, the specific loudness calculation is according to German standard DIN 45631/A1, which standardized the calculation of time-dependent loudness. The outcomes of the three different methods show similar results; therefore, the fluctuation of specific loudness is used for analysis of 1/f noise in critical bands in this study.

On the basis of the results of fluctuations of the parameters of loudness, sharpness, tonality, pitch, and specific loudness, respectively, 1/f noise behaviour is analysed using the computer program written by Dr. Bert De Coensel in Python, which has been used for the calculation of 1/f characteristic as a descriptor for temporal structure of soundscape (Botteldooren *et al.* 2006), and been extended by Dr. Bert De Coensel for statistical analyses of large amount of sound samples in this study.

3.4 Data Statistics

Based on the results of basic statistic indices of the parameters described above, a number of statistic methods are used with the software of SPSS Statistics, to explore the differences between natural and urban sounds and among the different categories, to examine the main dimensions of the indices, and to automatically identify/classify the sound categories.

3.4.1 One-way analysis of variance

To examine if the sound categories differ from each other significantly in one or more indices, one-way analysis of variance (ANOVA) is used to compare the means among the sound categories. The analysis of variance firstly tests the hypothesis that all group means are equal (F-test), and if a significant F-test suggests real differences among the means, it then involves a more detailed examination of the differences (Hilton and Armstrong 2006; Spss Inc. 2009).

3.4.2 Principal component analysis

Since it is noted that the parameters used for analysis (described in Section 3.3) have certain correlations, in order to reduce the dimensionality of the dataset and to determine the key influencing parameters, principal component analysis (PCA) is implemented. PCA transforms the large number of interrelated variables into a new set of uncorrelated variables, namely the principal components (PCs). The PCs are ordered in the way that the first component accounts for the most variance and the next one accounts for as much of the leftover variance as it can, and so on, so that the first few PCs retain most of the variation present in all of the original variables (Jolliffe 2002; UCLA Statistical Consulting Group 2013b).

3.4.3 Hierarchical cluster analysis

Hierarchical cluster analysis (HCA) attempts to identify relatively homogeneous groups of cases (or variables) based on selected characteristics, using an algorithm that starts with each case (or variable) in a separate cluster and combines clusters until only one is left (Spss Inc. 2009). HCA is used in this study to explore the grouping of the recordings – whether the clustered groups correspond to the categories of sound, that is, whether recordings in a given category show similarity in terms of any parameter. With each of the parameters, the procedure gradually clusters the recordings based on distance or similarity measures. At each step, it computes the distance between all cluster pairs and combines the two clusters with the smallest distance.

3.4.4 Discriminant function analysis

Discriminant function analysis (DFA), also known as discriminant analysis, is used to predict group membership (sound category) of the recordings with discriminant functions (DFs) (Spss Inc. 2009). From a sample of the cases for which group membership is known, the functions are generated, which are linear combinations of the variables or indices that provided the best discrimination between the categories. In other words, discriminant coefficients of the functions maximize the distance between the group means in the multidimensional space formed by the DFs. With discriminant scores, each recording is classified into a category by comparing distances to the group centroids, which are the mean discriminant scores for each category, or probabilities of group

membership. The functions can also be applied to new cases that have measurements for the predictor variables but have unknown group membership.

3.4.5 Artificial neural networks

In addition to DFA, artificial neural networks (ANNs) are developed to automatically identify the sound categories of recordings, i.e. water, wind, bird and urban sounds. ANN, inspired by and imitates the structure and function of biological nervous systems, is a computer learning system that makes predictions to questions from previous similar experiences it learnt (Kulkarni 1994). Unlike PCA and DFA, which analyse linearly, ANN can model nonlinear relationships. ANNs have been applied in various areas such as speech and visual image recognition (Patterson 1996), and the area of soundscape to predict reverberation times, sound levels and subjective evaluation of soundscape quality in urban open spaces (Yu 2003; Yu and Kang 2009).

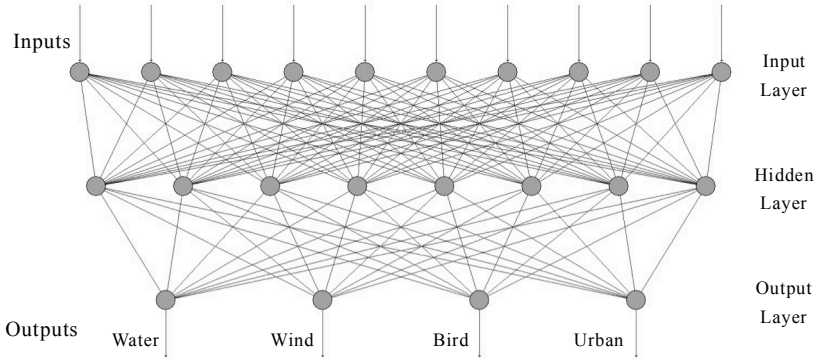


Figure 3.4.1 An example of artificial neural network architecture

It has been shown that different ANN software packages, including Qnet and NeuroSolutions, generally show similar prediction results (Yu and Kang 2009). Qnet is therefore used in this study. Qnet, which is a back-propagation style neural modelling system, requires supervised training, with a set of training data where known solutions are supplied (Vesta Services Inc. 2000). Processing elements (often termed neurons, units or nodes) contained in a network are organized in layers including input, hidden and output layers. Connections exist between the nodes of adjacent layers to relay signals. An example of the architecture of ANN is shown in Figure 3.4.1. The training algorithms iteratively adjust the nodal weights for the connections in an attempt to drive the network's response error, i.e. the differences between the responses at the output layer

and known answers supplied (training targets), to a minimum. Apart from the training set, test sets of data are not used to train the network, but monitor the network's responses to patterns outside the training set, to reflect network's overtraining status and to check the integrity of model.

3.5 Summary

This chapter described the methodology for this research, which aims to explore the differences between nature and urban sound and among different categories of sound in terms of objective measures.

First, natural sound in this research is defined as the sound that caused by nature and comes from the natural materials or objects. In other words, both the force causing the object's vibration and vibrating object are natural. Based on this definition, sounds used in this study, considering the most frequently heard sound sources in soundscapes of everyday life, are classified into natural sounds and all sounds that besides natural sounds, of which either or both the force and object is not natural, and is termed as urban sounds in this study. The natural sounds consist of three categories: Water sounds, wind sounds, and birdsongs; each category is further classified into several subcategories. The urban sounds, or human related sounds, include sounds of church bells, fountains, street music, street machines, traffic, human voice and footsteps.

Second, a large number of sound samples for the analysis in this study are collected from recordings made by the author and multiple databases. The recordings were made in countryside, natural parks, and urban areas in England, from 1994 to 2010. The absolute SPLs of the recordings were either recorded or in reasonable sound level ranges.

Third, the sound recordings in different categories are analysed in terms of objective parameters from three aspects, which are psychoacoustic parameters that have been recommended in previous soundscape research, additional psychoacoustic parameters that have mainly been applied in music perception, and dynamic indicator that has been used for analysing music and soundscapes.

Finally, characteristics of the four sound categories and differences among them are analysed using the statistic methods including one-way analysis of variance (ANOVA), principal component analysis (PCA), and hierarchical cluster analysis (HCA). Furthermore, categories of the recordings are automatically identified using discriminant function analysis (DFA) and artificial neural network (ANN).

In the following chapters, based on the calculated results of the sound samples collected in terms of the parameters in the three aspects, i.e., psychoacoustics, music, and dynamics, characteristics and identification of sound categories are analysed using generally all of the statistic methods in each of Chapters 4, 7, 8 and 9. While Chapters 4, 7 and 8 analyse the three aspects respectively, Chapter 9 analyses all the aspects together.

Chapter 4
Psychoacoustical analysis of natural and urban sounds in soundscapes

This chapter seeks to study the differences between natural and urban environmental sounds from the aspect of subjective sensations of the human hearing system, in particular relating to the field of psychoacoustics, since it can be expected that the human hearing system has been adapted to common natural sounds. Psychoacoustics studies the quantitative correlation between acoustical stimuli and hearing sensations. A number of psychoacoustic parameters that have been well developed are used to analyse the sound recordings in this chapter, which are loudness, sharpness, tonality, roughness and fluctuation strength, as well as sound pressure level (SPL).

In this chapter, Section 4.1 first examines possible differences between natural and urban environmental sounds and among different categories or subcategories of sounds, in terms of single parameters. Section 4.2 then analyses the differences based on all the parameters together, using principal component analysis (PCA). Finally, in Section 4.4, categories of the recordings are identified based on all or part of the parameters with artificial neural network (ANN) and discriminant function analysis (DFA).

4.1 Comparison among Various Types of Sound with Single Parameters

For each recording segment (see Chapter 3), the six parameters are calculated, i.e. the psychoacoustic parameters of loudness (N), sharpness (S), tonality (Ton), roughness (R) and fluctuation strength (Fls), and sound pressure level (L). From the time varying results, the statistic indices, which are average (AVE), standard deviation (STD), maximum (MAX) and minimum (MIN), over the time of 30 seconds of the six parameters are calculated for each segment. The results of these segments in terms of the indices are then averaged for each recording. The distributions of recordings in the 21 subcategories in the results of AVE indices of the six parameters are shown in Figure 4.1.1, in which the minimum, first quartile (25th percentile), median (50th percentile), third quartile (75th percentile) and maximum scores of each subcategory are presented. For recordings in each subcategory of sound, average values of the AVE and STD are

shown in Table 4.1.1. Based on these results, the differences among different categories or subcategories of sound in terms of single parameters are examined following with one-way analysis of variance (ANOVA) and hierarchical cluster analysis (HCA) in this section.

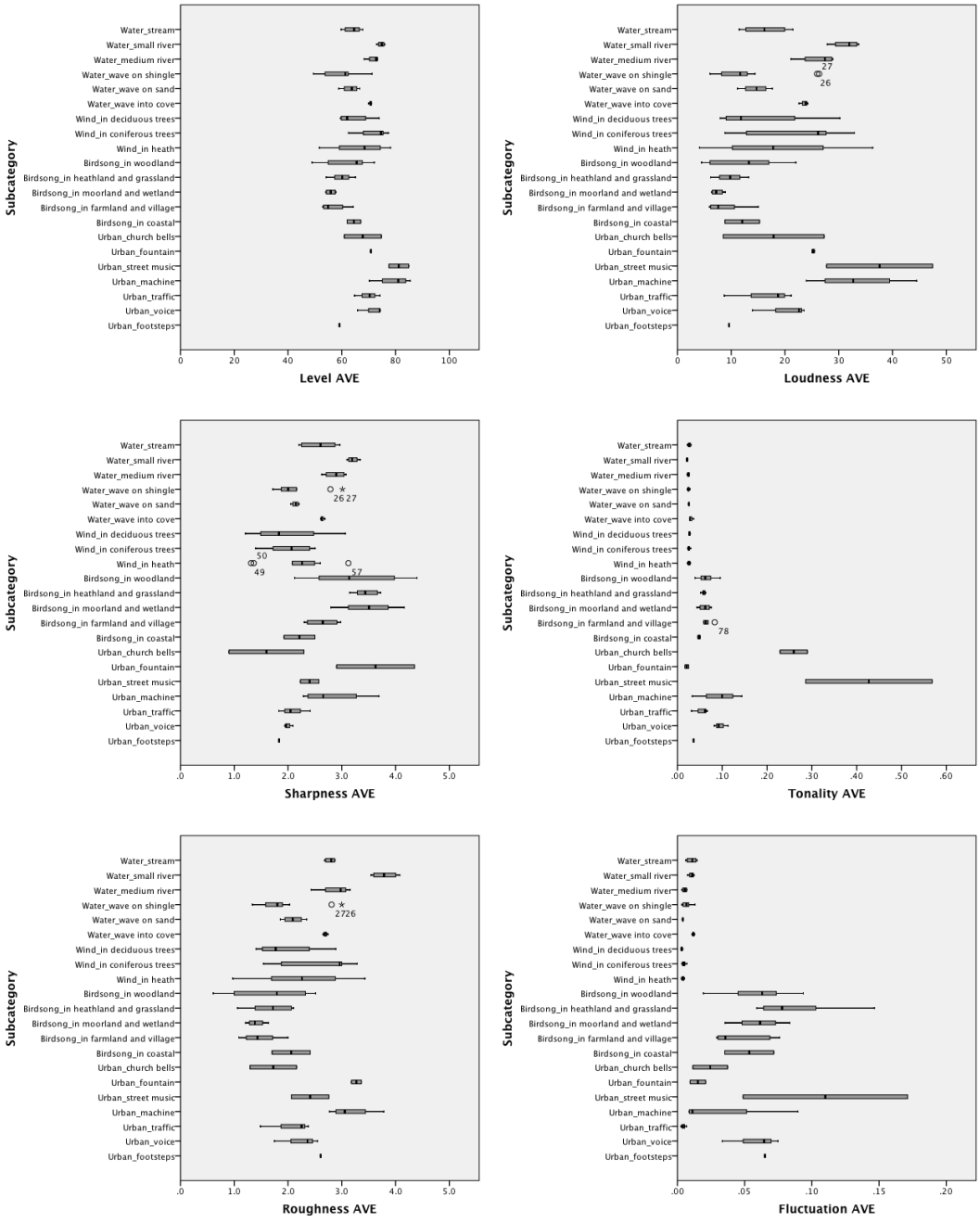


Figure 4.1.1 Statistic results of average values of SPL and psychoacoustic parameters of 21 subcategories of sound recordings, (a) level, (b) loudness, (c) sharpness, (d) tonality, (e) roughness, and (f) fluctuation strength

Table 4.1.1 Calculated results of SPL and psychoacoustic parameters for each sound subcategory

Category		Subcategory	SPL (dB)		Loudness (sone)		Sharpness (acum)		Tonality (tu)		Roughness (asper)		Fluctuation Strength (vacil)	
			AVE	STD	AVE	STD	AVE	STD	AVE	STD	AVE	STD	AVE	STD
Natural Sounds	Water Sounds	Stream	64.1	0.7	16.3	1.0	2.58	0.10	0.0263	0.0272	2.79	0.19	0.0107	0.0024
		Small river	74.7	0.6	31.4	1.7	3.21	0.12	0.0217	0.0251	3.80	0.22	0.0104	0.0026
		Medium river	71.8	0.3	26.2	0.8	2.88	0.05	0.0243	0.0277	2.89	0.10	0.0054	0.0011
		Wave on shingle	59.1	4.8	12.5	4.2	2.11	0.25	0.0249	0.0295	1.87	0.36	0.0072	0.0037
		Wave on sand	63.3	1.2	14.6	1.3	2.14	0.09	0.0255	0.0282	2.09	0.14	0.0040	0.0011
		Wave into cove	70.6	3.6	23.6	5.6	2.64	0.25	0.0303	0.0315	2.69	0.32	0.0118	0.0053
	Wind Sounds	Wind in deciduous trees	64.3	1.1	15.4	1.3	1.98	0.11	0.0268	0.0320	1.96	0.11	0.0032	0.0009
		Wind in coniferous trees	72.3	2.0	22.6	2.9	2.03	0.12	0.0257	0.0300	2.62	0.25	0.0049	0.0014
		Wind in heath	66.8	1.9	19.5	2.4	2.21	0.15	0.0260	0.0303	2.27	0.18	0.0041	0.0013
	Birdsongs	Birdsong in woodland	62.5	4.4	12.8	3.6	3.24	0.68	0.0639	0.0549	1.68	0.33	0.0587	0.0377
		Birdsong in heathland and grassland	59.9	6.4	9.7	4.1	3.45	0.79	0.0581	0.0492	1.68	0.56	0.0880	0.0625
		Birdsong in moorland and wetland	55.9	3.7	7.4	1.7	3.49	0.55	0.0609	0.0577	1.40	0.25	0.0605	0.0349
		Birdsong in farmland and village	56.7	4.1	8.8	2.6	2.64	0.62	0.0658	0.0603	1.48	0.30	0.0459	0.0332
		Birdsong in coastal	64.5	2.7	12.0	3.1	2.21	0.41	0.0482	0.0514	2.06	0.34	0.0534	0.0337
Urban Sounds		Church bells	67.8	4.6	17.9	6.6	1.60	0.23	0.2595	0.1855	1.73	0.22	0.0244	0.0153
		Fountain	70.8	0.7	25.3	1.7	3.63	0.15	0.0209	0.0248	3.27	0.24	0.0152	0.0047
		Street music	81.2	3.1	37.6	7.9	2.40	0.39	0.4272	0.1784	2.41	0.64	0.1100	0.0541
		Machine	79.4	3.0	33.5	6.2	2.82	0.29	0.0940	0.0544	3.17	0.36	0.0301	0.0167
		Traffic	69.8	3.4	16.2	4.0	2.10	0.12	0.0529	0.0485	2.04	0.27	0.0045	0.0026
		Voice	71.4	3.3	20.0	4.8	2.00	0.30	0.0957	0.0925	2.22	0.26	0.0576	0.0352
		Footsteps	59.2	3.2	9.6	2.6	1.83	0.27	0.0360	0.0460	2.61	1.29	0.0651	0.0596

4.1.1 Comparison among the four categories by the means of indices with one-way analysis of variance

To examine if the sound categories differ from each other significantly in one or more indices, one-way analysis of variance (ANOVA) is used to compare the mean values among the four categories in the software of SPSS Statistics, in terms of the 24 indices, i.e. AVE, STD, MAX and MIN of each of the six parameters (L, N, S, Ton, R and Fls). In this section, the analysis is based on the data of 102 segments – the first segment of each recording. The descriptives of the indices for the four categories are shown in Table 4.1.2, including mean, stand deviation, minimum, and maximum. While the ANOVA on AVE indices examines the differences between recordings in the absolute, mean value, the analysis on STD indices examines the differences in variation from the mean value.

The results of ANOVA are shown in Table 4.1.3, where the F ratio and p value, i.e. the significance of the F ratio (Sig.), are displayed. The F ratio shows the ratio of between-groups variance to within-groups variance, more specifically, is the ratio of mean square between-groups to that within-groups, in which the mean square is calculated by dividing the sum of square by its degrees of freedom (the between-groups degrees of freedom is the number of categories minus one, and for within-groups it is the sum of the number of cases in each category minus one) (Elvers 2013). From the table, it can be seen that S MIN and Ton MIN have p value greater than an alpha (α) level of 0.05, which suggests that there may be no statistically significant difference in the means of the indices among the four categories; for the other indices, the p value associated with the F ratio is less than 0.05, which reject the null hypothesis that all the means are equal. In other words, there may be some significant differences in terms of most the indices among the categories, or between at least two categories, whereas the assumption of homogeneity of variances and post hoc tests are further checked for verifying and examining which of the specific categories differ.

Test of homogeneity of variances examines the assumption of ANOVA that the variances of the categories are equal. Table 4.1.3 shows the results of Levene's test of homogeneity of variances of the indices. It can be seen that p values (Sig.) of L AVE, L MAX, L MIN, N MAX, S AVE and S MIN are greater than an α level of 0.05, which means that the variances are equal and the assumption of homogeneity of variances is met. For other indices, the p value (Sig.) are less than 0.05, thus the assumption is rejected, suggesting the variances are unequal. As a result, different post hoc tests based on the assumption of equal variances or not are executed.

Table 4.1.2 Descriptives of the psychoacoustic indices for the four categories

			Mean	Std. Deviation	Minimum	Maximum		Mean	Std. Deviation	Minimum	Maximum
L	Water	AVE	64.75	7.75	48.85	76.04	MAX	71.25	6.02	61.54	84.70
	Wind		68.66	8.27	51.15	78.39		73.77	8.75	58.51	86.92
	Bird		60.36	6.71	49.02	75.86		73.05	7.99	59.78	87.72
	Urban		73.21	7.48	60.92	87.02		82.12	7.18	71.03	91.35
	Water	STD	2.51	2.12	0.25	6.47	MIN	59.08	10.70	36.69	74.91
	Wind		1.95	1.16	0.49	4.59		63.98	10.90	26.94	74.31
	Bird		4.61	1.70	1.29	8.50		51.41	7.06	37.06	65.11
	Urban		3.11	1.97	0.30	7.72		67.11	7.28	54.21	83.07
N	Water	AVE	18.18	8.33	5.82	33.74	MAX	27.24	11.97	15.01	60.04
	Wind		20.04	10.62	3.63	33.23		28.26	16.78	5.18	59.49
	Bird		10.77	4.88	4.38	21.99		26.19	12.76	9.24	61.54
	Urban		24.58	11.11	8.51	48.84		44.11	19.32	10.24	87.85
	Water	STD	2.70	2.32	0.50	11.21	MIN	13.08	8.47	1.65	30.39
	Wind		2.92	2.77	0.40	10.80		14.74	8.25	0.28	26.33
	Bird		3.43	1.71	1.16	8.36		5.61	2.86	1.37	11.00
	Urban		5.11	3.23	0.31	11.34		15.51	6.42	5.59	28.67
S	Water	AVE	2.464	0.485	1.703	3.338	MAX	2.992	0.577	2.044	4.375
	Wind		2.080	0.528	1.157	3.033		2.684	0.762	1.466	4.310
	Bird		3.155	0.648	1.955	4.217		5.533	1.077	3.248	7.872
	Urban		2.412	0.765	0.903	4.337		3.549	1.010	1.248	5.026
	Water	STD	0.169	0.112	0.036	0.535	MIN	2.057	0.546	1.244	2.990
	Wind		0.146	0.101	0.052	0.442		1.781	0.493	0.803	2.627
	Bird		0.704	0.262	0.316	1.512		1.922	0.500	0.983	3.015
	Urban		0.239	0.120	0.046	0.435		1.944	0.710	0.782	4.055
Ton	Water	AVE	0.02512	0.00355	0.01937	0.03819	MAX	0.14214	0.02172	0.10456	0.18924
	Wind		0.02634	0.00260	0.02030	0.03245		0.16541	0.05001	0.12310	0.31904
	Bird		0.06379	0.02280	0.03894	0.14645		0.35186	0.11658	0.18374	0.67342
	Urban		0.13155	0.13322	0.01731	0.52668		0.41503	0.24070	0.09319	0.80541
	Water	STD	0.02840	0.00288	0.02388	0.03524	MIN	0.00000	0.00000	0.00000	0.00000
	Wind		0.03120	0.00425	0.02618	0.04237		0.00000	0.00000	0.00000	0.00000
	Bird		0.05558	0.02153	0.03215	0.12635		0.00000	0.00000	0.00000	0.00000
	Urban		0.08545	0.05939	0.02068	0.21000		0.00269	0.01109	0.00000	0.04571
R	Water	AVE	2.476	0.737	1.326	4.095	MAX	3.316	0.937	2.003	5.317
	Wind		2.347	0.836	0.722	3.335		3.029	1.198	0.898	5.108
	Bird		1.643	0.491	0.620	2.488		2.910	0.942	1.185	5.466
	Urban		2.592	0.745	1.293	4.021		4.453	2.552	1.647	12.533
	Water	STD	0.249	0.125	0.073	0.589	MIN	1.903	0.885	0.114	3.578
	Wind		0.210	0.146	0.065	0.568		1.908	0.742	0.141	2.779
	Bird		0.365	0.180	0.145	0.934		1.046	0.469	0.270	1.744
	Urban		0.413	0.432	0.054	1.903		1.796	0.638	0.937	2.828
Fls	Water	AVE	0.00797	0.00368	0.00339	0.01534	MAX	0.01655	0.00973	0.00501	0.03536
	Wind		0.00414	0.00133	0.00207	0.00739		0.00781	0.00410	0.00347	0.02012
	Bird		0.06207	0.03502	0.01355	0.16116		0.17595	0.10194	0.04326	0.43534
	Urban		0.03678	0.03509	0.00287	0.10777		0.10132	0.09841	0.00404	0.31311
	Water	STD	0.00271	0.00187	0.00059	0.00699	MIN	0.00397	0.00210	0.00126	0.00899
	Wind		0.00117	0.00084	0.00038	0.00390		0.00234	0.00074	0.00112	0.00345
	Bird		0.03696	0.02501	0.00530	0.12345		0.01558	0.01349	0.00218	0.05380
	Urban		0.02236	0.02311	0.00049	0.07580		0.00988	0.00911	0.00179	0.03107

Table 4.1.3 Test of homogeneity of variances and ANOVA of the psychoacoustic indices for the four categories

		Test of homogeneity of variances		ANOVA			Test of homogeneity of variances		ANOVA	
		Levene Statistic	Sig.	F	Sig.		Levene Statistic	Sig.	F	Sig.
L	AVE	0.70	0.552	11.62	0.000	MAX	2.61	0.056	8.44	0.000
	STD	4.88	0.003	10.88	0.000	MIN	1.83	0.147	12.49	0.000
N	AVE	7.22	0.000	10.10	0.000	MAX	2.32	0.080	6.32	0.001
	STD	2.84	0.042	3.96	0.010	MIN	10.22	0.000	10.74	0.000
S	AVE	0.99	0.402	15.02	0.000	MAX	3.52	0.018	61.87	0.000
	STD	7.39	0.000	68.65	0.000	MIN	0.17	0.916	1.15	0.334
Ton	AVE	28.42	0.000	16.51	0.000	MAX	35.09	0.000	31.45	0.000
	STD	30.46	0.000	20.93	0.000	MIN	7.74	0.000	1.70	0.172
R	AVE	3.48	0.019	9.57	0.000	MAX	5.49	0.002	4.93	0.003
	STD	4.24	0.007	4.12	0.009	MIN	3.22	0.026	9.26	0.000
Fls	AVE	25.10	0.000	37.02	0.000	MAX	28.42	0.000	38.43	0.000
	STD	22.64	0.000	30.60	0.000	MIN	20.32	0.000	15.08	0.000

Post hoc tests are executed to examine which of the means for the four categories significantly differ from the others, since significant differences exist among the means of the groups. A variety of methods are available for conducting post hoc tests. The most common tests that assume homogeneity of variances include least significant difference (LSD) test – the original solution developed by Fisher which uses t tests to perform all pairwise comparisons between group means –, Tukey's honestly significant difference (Tukey's HSD) test, which was developed in reaction to the LSD test and uses the more conservative Studentized range statistic to make all pairwise comparisons between groups, the Student-Newman-Keuls (SNK) method, which uses the Studentized range distribution to make all pairwise comparisons with stepwise procedure available, Bonferroni method, which uses t tests to perform pairwise comparisons between group means, but adjusts the error rate for each test in multiple comparisons, and Scheffé's test, which performs simultaneous joint pairwise comparisons using the F sampling distribution (Stevens 1999; Spss Inc. 2009). Here, Tukey's HSD and Bonferroni methods are used for the indices with equal variances between categories, considering the unequal sample sizes in groups.

The post hoc tests that do not assume homogeneity of variances include Tamhane's T2, Dunnett's T3, Games - Howell and Dunnett's C. While Tamhane's T2, T3 and C are conservative pairwise comparisons tests based on t test, Studentized maximum modulus, and Studentized range respectively, Games and Howell test is an extension of Tukey-Kramer test to the case of unequal variances (Cardinal 2013). As for the indices (with unequal variances) the four methods generally generate the same results, the results of Dunnett's T3 test are presented here.

Table 4.1.4 Multiple comparisons for the four categories of the psychoacoustic indices (* indicates that the mean difference is significant at the 0.05 level)

Mean Difference (I-J)		(I) Category	Water			Wind			Bird			Urban		
Dependent Variable		(J) Category	Wind	Bird	Urban	Water	Bird	Urban	Water	Wind	Urban	Water	Wind	Bird
L	AVE	Tukey HSD	-3.92	4.39	-8.47*	3.92	8.31*	-4.55	-4.39	-8.31*	-12.86*	8.47*	4.55	12.86*
	STD	Dunnett T3	0.56	-2.10*	-0.60	-0.56	-2.66*	-1.16	2.10*	2.66*	1.50	0.60	1.16	-1.50
	MAX	Tukey HSD	-2.52	-1.79	-10.87*	2.52	0.72	-8.35*	1.79	-0.72	-9.07*	10.87*	8.35*	9.07*
	MIN	Tukey HSD	-4.90	7.67*	-8.03*	4.90	12.57*	-3.13	-7.67*	-12.57*	-15.70*	8.03*	3.13	15.70*
N	AVE	Dunnett T3	-1.86	7.41*	-6.40	1.86	9.27*	-4.54	-7.41*	-9.27*	-13.81*	6.40	4.54	13.81*
	STD	Dunnett T3	-0.22	-0.73	-2.41	0.22	-0.51	-2.20	0.73	0.51	-1.68	2.41	2.20	1.68
	MAX	Tukey HSD	-1.02	1.05	-16.87*	1.02	2.06	-15.85*	-1.05	-2.06	-17.91*	16.87*	15.85*	17.91*
	MIN	Dunnett T3	-1.66	7.46*	-2.43	1.66	9.13*	-0.77	-7.46*	-9.13*	-9.90*	2.43	0.77	9.90*
S	AVE	Tukey HSD	0.384	-0.691*	0.052	-0.384	-1.075*	-0.332	0.691*	1.075*	0.743*	-0.052	0.332	-0.743*
	STD	Dunnett T3	0.024	-0.535*	-0.069	-0.024	-0.558*	-0.093	0.535*	0.558*	0.465*	0.069	0.093	-0.465*
	MAX	Dunnett T3	0.308	-2.541*	-0.557	-0.308	-2.849*	-0.866*	2.541*	2.849*	1.984*	0.557	0.866*	-1.984*
	MIN	Tukey HSD	0.276	0.135	0.113	-0.276	-0.141	-0.162	-0.135	0.141	-0.021	-0.113	0.162	0.021
Ton	AVE	Dunnett T3	-0.0012	-.0387*	-.1064*	0.0012	-.0375*	-.1052*	.0387*	.0375*	-0.0678	.1064*	.1052*	0.0678
	STD	Dunnett T3	-0.0028	-.0272*	-.0570*	0.0028	-.0244*	-.0543*	.0272*	.0244*	-0.0299	.0570*	.0543*	0.0299
	MAX	Dunnett T3	-0.0233	-.2097*	-.2729*	0.0233	-.1865*	-.2496*	.2097*	.1865*	-0.0632	.2729*	.2496*	0.0632
	MIN	Dunnett T3	0.0000	0.0000	-0.0027	0.0000	0.0000	-0.0027	0.0000	0.0000	-0.0027	0.0027	0.0027	0.0027
R	AVE	Dunnett T3	0.129	0.834*	-0.115	-0.129	0.705*	-0.244	-0.834*	-0.705*	-0.949*	0.115	0.244	0.949*
	STD	Dunnett T3	0.039	-0.116*	-0.164	-0.039	-0.155*	-0.203	0.116*	0.155*	-0.048	0.164	0.203	0.048
	MAX	Dunnett T3	0.287	0.406	-1.138	-0.287	0.119	-1.425	-0.406	-0.119	-1.543	1.138	1.425	1.543
	MIN	Dunnett T3	-0.005	0.857*	0.107	0.005	0.863*	0.112	-0.857*	-0.863*	-0.751*	-0.107	-0.112	0.751*
Fls	AVE	Dunnett T3	0.0038*	-.0541*	-.0288*	-.0038*	-.0579*	-.0326*	.0541*	.0579*	0.0253	0.0288*	.0326*	-0.0253
	STD	Dunnett T3	0.0015*	-.0343*	-.0197*	-.0015*	-.0358*	-.0212*	.0343*	.0358*	0.0146	0.0197*	.0212*	-0.0146
	MAX	Dunnett T3	0.0087*	-.1594*	-.0848*	-.0087*	-.1681*	-.0935*	.1594*	.1681*	0.0746	0.0848*	.0935*	-0.0746
	MIN	Dunnett T3	0.0016*	-.0116*	-0.0059	-.0016*	-.0132*	-.0075*	.0116*	.0132*	0.0057	0.0059	.0075*	-0.0057
L	AVE	Bonferroni	-3.92	4.39	-8.47*	3.92	8.31*	-4.55	-4.39	-8.31*	-12.86*	8.47*	4.55	12.86*
	STD		0.56	-2.10*	-0.60	-0.56	-2.66*	-1.16	2.10*	2.66*	1.50*	0.60	1.16	-1.50*
	MAX		-2.52	-1.79	-10.87*	2.52	0.72	-8.35*	1.79	-0.72	-9.07*	10.87*	8.35*	9.07*
	MIN		-4.90	7.67*	-8.03*	4.90	12.57*	-3.13	-7.67*	-12.57*	-15.70*	8.03*	3.13	15.70*
N	AVE	Bonferroni	-1.86	7.41*	-6.40	1.86	9.27*	-4.54	-7.41*	-9.27*	-13.81*	6.40	4.54	13.81*
	STD		-0.22	-0.73	-2.41*	0.22	-0.51	-2.20*	0.73	0.51	-1.68	2.41*	2.20*	1.68
	MAX		-1.02	1.05	-16.87*	1.02	2.06	-15.85*	-1.05	-2.06	-17.91*	16.87*	15.85*	17.91*
	MIN		-1.66	7.46*	-2.43	1.66	9.13*	-0.77	-7.46*	-9.13*	-9.90*	2.43	0.77	9.90*
S	AVE	Bonferroni	0.384	-.691*	0.052	-0.384	-1.075*	-0.332	.691*	1.075*	.743*	-0.052	0.332	-.743*
	STD		0.024	-.535*	-0.069	-0.024	-.558*	-0.093	.535*	.558*	.465*	0.069	0.093	-.465*
	MAX		0.308	-2.541*	-0.557	-0.308	-2.849*	-.866*	2.541*	2.849*	1.984*	0.557	.866*	-1.984*
	MIN		0.276	0.135	0.113	-0.276	-0.141	-0.162	-0.135	0.141	-0.021	-0.113	0.162	0.021
Ton	AVE	Bonferroni	-0.0012	-.0387*	-.1064*	0.0012	-0.0375	-.1052*	.0387*	0.0375	-.0678*	.1064*	.1052*	.0678*
	STD		-0.0028	-.0272*	-.0570*	0.0028	-.0244*	-.0543*	.0272*	.0244*	-.0299*	.0570*	.0543*	.0299*
	MAX		-0.0233	-.2097*	-.2729*	0.0233	-.1865*	-.2496*	.2097*	.1865*	-0.0632	.2729*	.2496*	0.0632
	MIN		0.0000	0.0000	-0.0027	0.0000	0.0000	-0.0027	0.0000	0.0000	-0.0027	0.0027	0.0027	0.0027
R	AVE	Bonferroni	0.129	.834*	-0.115	-0.129	.705*	-0.244	-.834*	-.705*	-.949*	0.115	0.244	.949*
	STD		0.039	-0.116	-0.164	-0.039	-0.155	-.203*	0.116	0.155	-0.048	0.164	.203*	0.048
	MAX		0.287	0.406	-1.138*	-0.287	0.119	-1.425*	-0.406	-0.119	-1.543*	1.138*	1.425*	1.543*
	MIN		-0.005	.857*	0.107	0.005	.863*	0.112	-.857*	-.863*	-.751*	-0.107	-0.112	.751*
Fls	AVE	Bonferroni	0.0038	-.0541*	-.0288*	-0.0038	-.0579*	-.0326*	.0541*	.0579*	.0253*	.0288*	.0326*	-.0253*
	STD		0.0015	-.0343*	-.0197*	-0.0015	-.0358*	-.0212*	.0343*	.0358*	.0146*	.0197*	.0212*	-.0146*
	MAX		0.0087	-.1594*	-.0848*	-0.0087	-.1681*	-.0935*	.1594*	.1681*	.0746*	.0848*	.0935*	-.0746*
	MIN		0.0016	-.0116*	-0.0059	-0.0016	-.0132*	-.0075*	.0116*	.0132*	0.0057	0.0059	.0075*	-0.0057

The results of the post hoc tests are shown in Table 4.1.4, by either Tukey's HSD or Dunnett's T3 method according to the homogeneity of variances of the categories in the

index and Bonferroni method for all the indices. It shows the multiple comparisons among the categories, in terms of difference between the means of each category with each of the other three, where asterisks indicate significantly different group means at an alpha level of 0.05 (Spss Inc. 2009). For the ones that p value is less than or equal to the α level of 0.05 (indicated by *), the corresponding null hypothesis that the means are equal are rejected, which means that there is a significant difference between the two categories in the index, while for the ones that p value is larger than 0.05, there may be no differences between the categories. Between the methods of Tukey's HSD and Bonferroni, generally, no difference in the results has been shown, thus, for indices with equal variances of the categories, only the results by Tukey's HSD method are discussed following.

From Table 4.1.4, it can be seen that in general there are significant differences between at least two categories, in terms of the indices except for N STD, S MIN, Ton MIN, and R MAX. The specific differences are discussed briefly here with the AVE indices. The index of L AVE shows significant differences between categories of water and urban, between wind and bird, and between bird and urban. The group mean value of L AVE of urban sound category is larger than that of wind, than water, and than bird. N AVE shows significant differences between bird and the other three categories. Birdsongs have lower mean value of N AVE than the other three. S AVE also shows significant differences between bird category and the other three, while birdsongs have higher mean value of S AVE than the other three. Ton AVE shows significant differences between the set of categories of water and wind and the set of bird and urban. Group mean values of Ton AVE of water and wind sounds are lower than those of bird and urban. R AVE shows significant differences between bird and the other three categories. Birdsongs have lower mean value of R AVE than the other three. Fls AVE shows significant differences between each pair of categories, except for bird and urban. Group mean values of Fls AVE of bird and urban sounds are larger that of water, and than wind.

In other words, there are significant differences among the categories in the majority of the indices. In terms of average values, birdsongs have higher sharpness, and lower loudness and roughness than the other three types of sounds. Also, bird and urban sounds have higher tonality and fluctuation strength than water and wind sounds. Differences among the categories and subcategories are further discussed in the next section (Section 4.1.2).

4.1.2 Hierarchical cluster analysis based on single parameters

In addition to ANOVA, the differences between natural and urban sounds and among the different categories with any single parameter are explored with hierarchical cluster analysis (HCA). The procedure of hierarchical cluster attempted to identify relatively homogeneous groups of the 102 recordings based on AVE, STD, MAX and MIN of each of the parameters, using the method of average linkage between groups in the software of SPSS Statistics 20. Here, for each recording, the averaged results of the segments are used.

With each of the parameters, the 102 recordings were clustered gradually using hierarchical cluster analysis that starts with each case in a separate cluster and then combines clusters until only one is left. The dendrograms for each of the parameters are shown in Figure 4.1.2. It has been shown that, compared to SPL, the differences among various types of sound are more significant in terms of loudness. The categories are rather mixed in the clusters based on SPL, whereas for loudness, one of the last two clusters has street music and machine sounds in the urban category, two sounds in water category and one sound in wind category. Street music and machine sounds generally have highest loudness among subcategories, with an average value of nearly 40 sone, as shown in Table 4.1.1. Birdsongs and footsteps have the lowest average values, at around 10 sone. The loudness of other types of sound is similar, in the range of around 20 sone. It is also noted that the standard deviations among the recordings in some subcategories, including wind, church bells, music, and machines, are relatively high, at about 10 to 20 sone (as shown in Figure 4.1.1). Generally speaking, these results show a tendency that urban sounds are slightly louder than natural sounds, or in other words, sounds with higher loudness are mainly urban sounds.

For sharpness, one of the last three clusters has most of the birdsongs and some of the music and machine sounds. Small rivers, most subcategories of birdsongs, fountains and machine sounds have relatively high average values of over 3.0 acum, while the other sounds are in the range of about 1.8 to 3.0 acum, except for church bells, with an average sharpness of about 1.6 acum. The standard deviations among the recordings in some subcategories, including birdsongs in woodland, fountains, church bells and machine sounds, are relatively high, at about 1.5 acum.

Based on the tonality average values, the subcategories are ordered from the highest to the lowest as: music (0.43 tu), church bells (0.26 tu), voice (0.10 tu), machine sounds (0.09 tu), birdsongs and traffic (both about 0.05 to 0.06 tu), water, wind, fountains and footsteps (about 0.02 to 0.03 tu). One of the last clusters has music, church bells and one of the voice recordings. The results show that generally the sounds with high tonality are

from urban sounds, while those with low values could be from both natural and urban sounds. Although the tonality algorithm is developed to calculate tonal contribution to euphony of sound, especially for music, for environmental sounds it may not correspond to people's preference of various sound types.

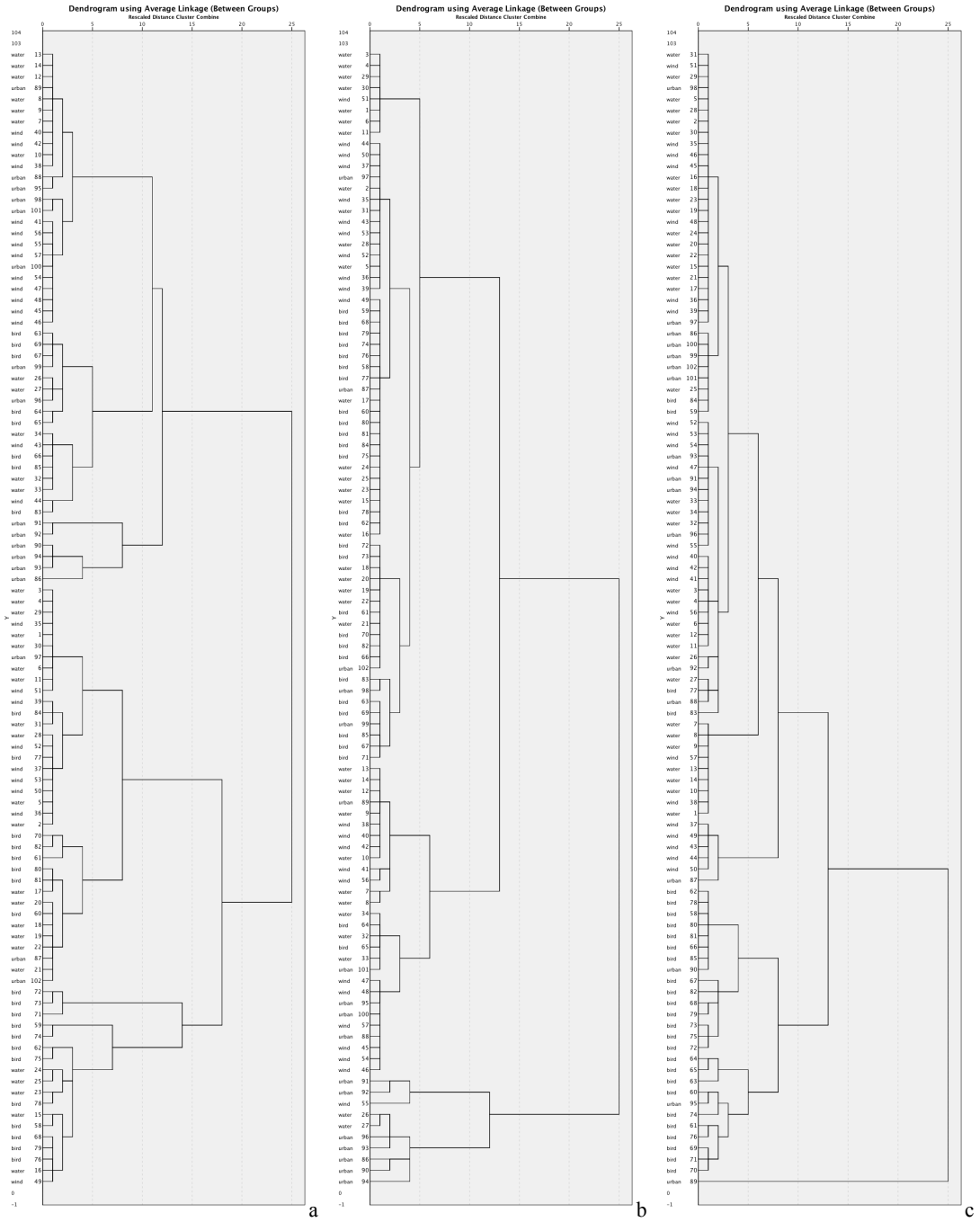


Figure 4.1.2 Dendrograms for the 102 recordings based on (a) level, (b) loudness, and (c) sharpness by HCAs

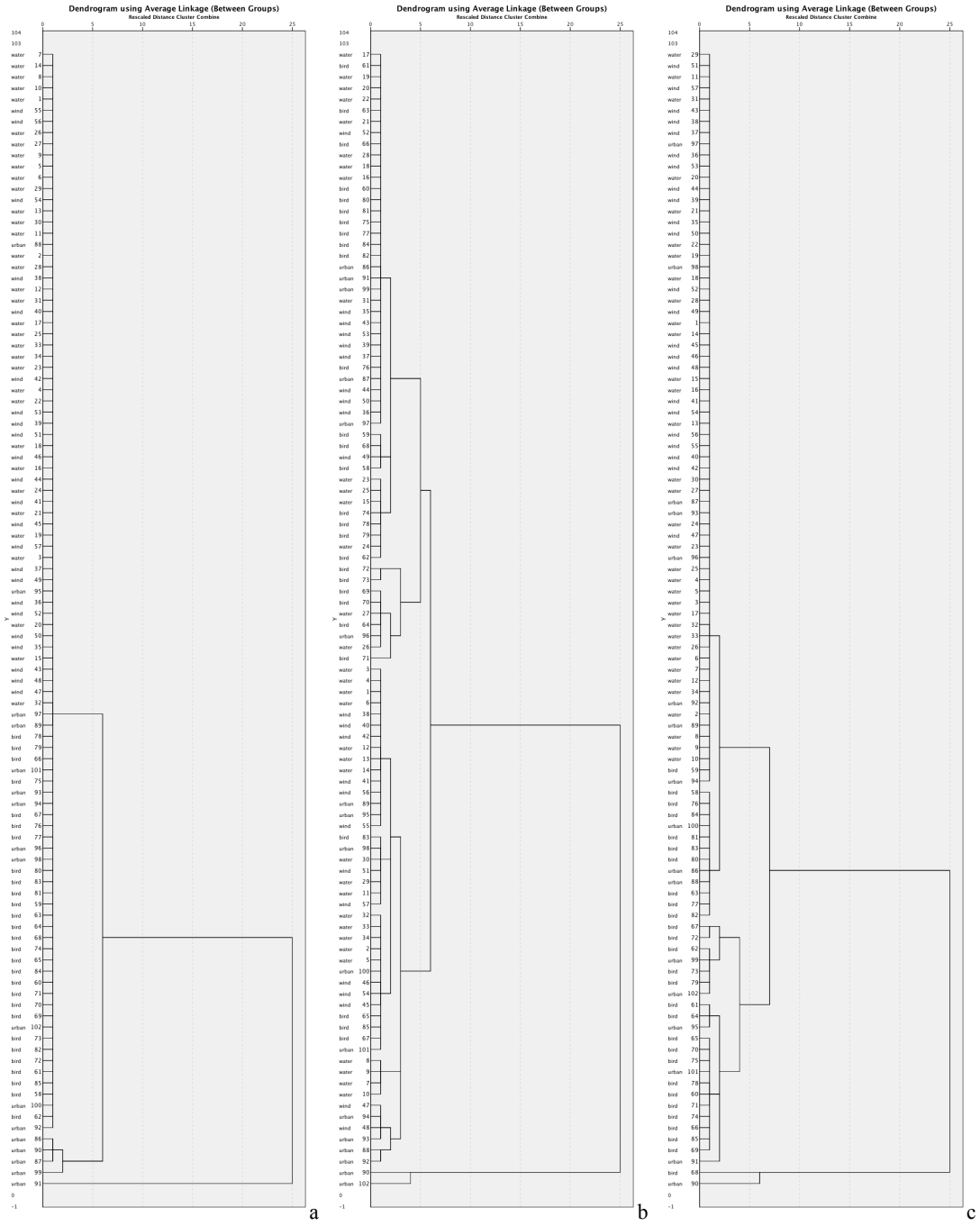


Figure 4.1.3 Dendrograms for the 102 recordings based on (a) tonality, (b) roughness, and (c) fluctuation strength by HCAs

The results of roughness show that small rivers, fountains and machine sounds have the highest average roughness values among all the subcategories, at over 3.0 asper; while birdsongs and church bells have the lowest values of around 1.5 asper. This provides an order of sound types based on auditory sensation from rough to smooth, although it is

hard to link this order to people's preference directly or to distinguish natural and urban sounds. In terms of clustering based on roughness, the various types of sound are rather mixed.

The fluctuation strength of birdsongs, music, human voice and footsteps, with average values of about 0.04 to 0.11 vacil, are much higher than those of water, wind, fountains and traffic (about 0.004 to 0.012 vacil), while church bells and machine sounds have the medium average values, at about 0.02 to 0.03 vacil. One of the last clusters in terms of fluctuation strength contains the majority of birdsongs and some of the machine sounds, music, human voice and footsteps, and another one contains sounds such as water, wind and traffic.

Overall, from the above results it can be seen that, although some sound types or subcategories are clustered together in terms of certain parameters like sharpness and fluctuation strength, and some parameters such as loudness and tonality show certain tendencies for natural and urban sounds, none of these parameters alone can be used to simply distinguish natural and urban sounds.

4.2 Principal Components Analysis

As any single parameter cannot distinguish natural and urban sounds, all these parameters are considered simultaneously to explore the differences between them. It is noted that these six parameters, or the 24 indices when considering AVE, STD, MAX and MIN of each parameter, have certain correlations. For example, loudness, sharpness and roughness to some degrees depend on SPL. In order to reduce the dimensionality of the dataset and to determine the key influencing parameters, principal component analysis (PCA) is implemented with software of SPSS Statistics 20, which, based on the psychoacoustical results of the 102 recordings, transforms the large number of interrelated variables into a smaller number of uncorrelated variables, namely the principal components, while retaining most of the variation present in the original variables (Jolliffe 2002). In this section, the data of the first segment of each of the 102 recordings are used.

In Section 4.2.1, the correlations between the 24 indices are examined. In Section 4.2.2, PCAs are carried out to extract the principal components of the indices and to find out key indices, based on different sets of indices respectively. Consequently, in Section 4.3, particular characteristics of each category and differences among the categories can be shown.

4.2.1 Correlations between the psychoacoustic indices

The correlations between the 24 indices, i.e. AVE, STD, MAX and MIN of sound pressure level (L), loudness (N), sharpness (S), tonality (Ton), roughness (R), and fluctuation strength (Fls), are examined before carrying out the PCA. The results are shown in Table 4.2.1. It can be seen that for each of the six parameters, the MAX and MIN indices are highly related to the AVE ones; and the MAX and STD indices are also highly related except for L, the MAX of which is highly related to N STD. These results suggest that the MAX and MIN indices carry similar information to that of the AVE and STD ones, or in other words, the AVE and STD indices alone generally cover most of information. Among the AVE and STD of the six parameters, the correlations among L AVE, N AVE and R AVE, among S STD, Fls AVE and Fls STD, and between Ton AVE and Ton STD are high, all of which are above about 0.7. These suggest that loudness and roughness somewhat relate to SPL, and both fluctuation strength and STD of certain parameters reflect the variation of sound.

4.2.2 Principal components of the psychoacoustic indices

The principal components of the indices are analysed using PCA in SPSS Statistics, based on the psychoacoustic results of the 102 recordings. The analysis is firstly executed based on the average data, and then based on the average and standard deviation data, and based on all the average, standard deviation, maximum and minimum data of the six parameters. Additionally, it is executed based on only the average and standard deviation data of the five psychoacoustic parameters, without SPL data.

Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is conducted before the PCAs, the results of which are shown in Table 4.2.2. For PCA based on the AVE data, the adequacy value is not high (about 0.5, while a value of 0.6 is a suggested minimum (UCLA Statistical Consulting Group 2013b)), thus the PCA based on the average indices is not presented here.

Table 4.2.1 Correlation matrix for AVE, STD, MAX, and MIN indices of SPL and psychoacoustic parameters

	L AVE	N AVE	S AVE	Ton AVE	R AVE	Fls AVE	L STD	N STD	S STD	Ton STD	R STD	Fls STD	L MAX	L MIN	N MAX	N MIN	S MAX	S MIN	Ton MAX	Ton MIN	R MAX	R MIN	Fls MAX	Fls MIN
L AVE	1.000																							
N AVE	**0.941**	1.000																						
S AVE	0.172	0.219	1.000																					
Ton AVE	0.240	0.217	-0.019	1.000																				
R AVE	**0.809**	**0.862**	0.244	-0.119	1.000																			
Fls AVE	-0.123	-0.199	0.446	0.304	-0.177	1.000																		
L STD	-0.336	-0.343	0.159	0.208	-0.445	0.434	1.000																	
N STD	0.431	0.410	0.126	0.385	0.227	0.227	0.588	1.000																
S STD	-0.187	-0.225	0.551	0.215	-0.308	**0.744**	0.663	0.390	1.000															
Ton STD	0.117	0.044	-0.046	**0.825**	-0.203	0.441	0.309	0.410	0.334	1.000														
R STD	0.032	0.026	0.139	0.150	0.154	0.486	0.533	0.544	0.420	0.219	1.000													
Fls STD	-0.155	-0.220	0.325	0.242	-0.209	**0.898**	0.486	0.247	0.694	0.371	0.497	1.000												
L MAX	**0.803**	**0.730**	0.231	0.376	0.570	0.167	0.218	**0.760**	0.190	0.317	0.383	0.189	1.000											
L MIN	**0.928**	**0.884**	0.079	0.130	**0.817**	-0.233	-0.571	0.198	-0.359	0.006	-0.128	-0.245	0.620	1.000										
N MAX	**0.760**	**0.785**	0.234	0.348	0.645	0.117	0.135	**0.784**	0.134	0.294	0.408	0.157	**0.899**	0.608	1.000									
N MIN	**0.849**	**0.914**	0.190	0.052	**0.865**	-0.302	-0.588	0.078	-0.400	-0.103	-0.171	-0.317	0.519	**0.910**	0.559	1.000								
S MAX	0.004	-0.016	**0.813**	0.155	-0.030	**0.703**	0.483	0.326	**0.855**	0.235	0.390	0.650	0.316	-0.139	0.296	-0.135	1.000							
S MIN	0.427	0.478	**0.736**	-0.175	0.578	-0.040	-0.334	-0.079	-0.089	-0.274	-0.127	-0.116	0.226	0.470	0.255	0.597	0.337	1.000						
Ton MAX	0.039	-0.049	0.062	**0.800**	-0.301	0.559	0.423	0.411	0.497	**0.903**	0.255	0.474	0.330	-0.091	0.262	-0.212	0.384	-0.296	1.000					
Ton MIN	0.199	0.239	-0.030	**0.709**	-0.033	0.019	-0.010	0.126	-0.022	0.228	-0.031	-0.029	0.209	0.157	0.182	0.167	-0.053	-0.041	0.341	1.000				
R MAX	0.523	0.541	0.219	0.062	**0.729**	0.227	-0.030	0.406	0.028	0.068	**0.736**	0.231	0.571	0.445	0.661	0.443	0.233	0.294	0.018	-0.045	1.000			
R MIN	**0.770**	**0.797**	0.208	-0.215	**0.918**	-0.336	-0.657	-0.008	-0.429	-0.309	-0.191	-0.360	0.406	**0.863**	0.455	**0.907**	-0.138	0.645	-0.400	-0.051	0.467	1.000		
Fls MAX	-0.149	-0.215	0.378	0.283	-0.223	**0.903**	0.497	0.263	**0.723**	0.422	0.466	**0.972**	0.208	-0.250	0.186	-0.325	**0.713**	-0.078	0.535	-0.021	0.204	-0.371	1.000	
Fls MIN	0.041	-0.038	0.473	0.307	-0.010	**0.817**	0.206	0.183	0.591	0.398	0.329	0.521	0.187	-0.066	0.146	-0.133	0.598	0.110	0.509	0.084	0.214	-0.128	0.588	1.000

4.2.2.1 Principal components analysis based on average and standard deviation indices

With the AVE and STD data of the six parameters for the 102 recordings, the PCA is conducted on the correlation matrix of the 12 indices (can be seen in Table 4.2.1). The adequacy value by KMO test is above 0.6, as shown in Table 4.2.2, which suggests the sample size is generally adequate. From the 12 indices or variables, 12 components are extracted, the initial number of which is the same as the number of variables used. Table 4.2.3 shows the variances explained by the components, in terms of eigenvalue, i.e. the variance of component, the percentage of variance explained by each component, and the cumulative percentage of variance explained by the current and all preceding components. The first component accounts for the most variance and hence has the highest eigenvalue; and the next one accounts for as much of the leftover variance as it can, and so on (UCLA Statistical Consulting Group 2013b). As each variable has a variance of 1 (so the total variance is 12 in this case), the first four principal components with eigenvalues greater than 1 are retained, which account for more variance than did the original variables. The first three components together account for 76.6% of the total variance, while the first four components together account for 86.2% of the total variance.

Table 4.2.2 KMO tests for AVE data, for AVE and STD data, for AVE, STD, MAX, and MIN data of the six parameters, and for AVE and STD data of the five psychoacoustic parameters

	AVE of the 6 parameters	AVE and STD of the 6 parameters	AVE, STD, MAX, and MIN of the 6 parameters	AVE and STD of the 5 parameters
KMO Measure of Sampling Adequacy	0.503	0.668	0.693	0.560

Table 4.2.3 Total variance explained by components based on AVE and STD indices, and on AVE, STD, MAX, and MIN indices of the six parameters

	AVE and STD of the 6 parameters			AVE, STD, MAX, and MIN of the 6 parameters		
Component	Eigenvalues	% of Variance	Cumulative %	Eigenvalues	% of Variance	Cumulative %
1	**4.346**	**36.218**	**36.218**	**7.806**	**32.527**	**32.527**
2	**3.175**	**26.455**	**62.673**	**7.143**	**29.761**	**62.288**
3	**1.671**	**13.925**	**76.597**	**2.978**	**12.407**	**74.695**
4	**1.152**	**9.602**	**86.200**	**1.931**	**8.046**	**82.741**
5	0.742	6.187	92.387	**1.219**	**5.078**	**87.818**
6	0.339	2.824	95.210	.829	3.452	91.271
7	0.178	1.486	96.696	.615	2.561	93.831
8	0.154	1.280	97.977	.471	1.964	95.796
9	0.084	0.703	98.680	.255	1.063	96.858
10	0.069	0.572	99.252	.166	.693	97.552
11	0.061	0.512	99.764	.142	.593	98.144
12	0.028	0.236	100.000	.119	.497	98.641

Table 4.2.4 Component matrix and communalities for AVE and STD indices of the six parameters

	Component Matrix				Communalities	
	Component 1	Component 2	Component 3	Component 4	Extraction of 3 components	Extraction of 4 components
L AVE	-.222	**.936**	-.023	.076	.926	.932
N AVE	-.275	**.941**	.024	.043	.962	.964
S AVE	.352	.296	**.650**	.357	.634	.762
Ton AVE	.449	.356	**-.696**	.291	.813	.898
R AVE	-.369	**.825**	.321	-.038	.919	.920
Fls AVE	**.850**	.037	.219	.330	.772	.881
L STD	**.788**	-.143	-.030	-.474	.642	.867
N STD	.507	**.603**	-.160	-.497	.647	.893
S STD	**.864**	-.021	.271	.088	.820	.828
Ton STD	.583	.245	**-.664**	.257	.841	.906
R STD	**.615**	.269	.165	-.482	**.479**	.711
Fls STD	**.833**	-.009	.209	.210	.738	.782

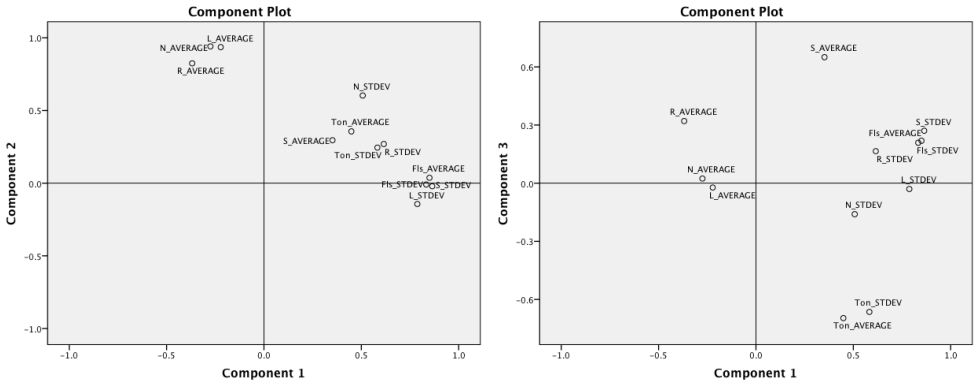


Figure 4.2.1 Loading plot of the principal components based on AVE and STD data

Table 4.2.4 shows component matrix of component loadings, i.e. the correlations between the principal components and the variables or indices. It shows that the correlations between Component 1 and Fls AVE, L STD, S STD, R STD and Fls STD, between Component 2 and L AVE, N AVE, R AVE, N STD, and between Component 3 and S AVE, Ton AVE and Ton STD are high (above 0.6). Component 4 has no particularly high correlation with any of the indices. These results can also be seen on the component loading plots, as shown in Figure 4.2.1, where only the first three components are displayed. It can be seen that some of the indices are clustered in groups, e.g. Fls AVE, L STD, S STD and Fls STD; and L AVE, N AVE and R AVE, where the indices in each group generally have high correlations among them, as shown in Table 4.2.1. Among these, S STD, N AVE and Ton AVE have the highest correlations with Components 1, 2 and 3 respectively. It is clear that Component 1 mainly represents fluctuation related

properties; Component 2 mainly represents loudness related properties; and Component 3 mainly represents tonality properties.

Table 4.2.4 also shows the proportion of each index's variance that can be explained by the retained principal components. When three components are retained from the 12 components, all the indices have relatively high proportion values except for R STD, which has a value below 0.5. When four components are retained, the proportions of all the indices' variance that can be explained are high, which means that all these indices are well represented by the principal components.

4.2.2.2 Principal components analysis based on average, standard deviation, maximum, and minimum indices

Additionally, PCA is carried out based on all the AVE, STD, MAX and MIN data. In Table 4.2.2, it shows that the adequacy value of KMO test is above 0.6, which suggests the sample size is generally adequate. From the 24 variables, 24 components are obtained. Table 4.2.3 shows the variances explained by the first 12 components. It can be seen that the eigenvalues of the first five components are greater than 1. The first three components together account for 74.7% of the total variance, while the first four components together account for 82.7% and the first five components account for 87.8% of the total variance.

Table 4.2.5 shows the correlations between the first five principal components and the indices and the proportion of each index's variance that can be explained by the retained principal components. It shows that Component 1 has high correlations (above 0.6) with L AVE, N AVE, R AVE, LSTD, S STD, L MIN, N MIN and R MIN; Component 2 has high correlations with Fls AVE, N STD, Fls STD, L MAX, N MAX, S MAX, R MAX, Fls MAX and Fls MIN; Component 3 has high correlations with S AVE and Ton AVE; Components 4 and 5 have no particularly high correlation with any of the variables. It suggests that Component 1 mainly represents loudness proprieties, including AVE and MIN indices of L, N and R; Component 2 mainly represents fluctuation proprieties, including AVE, STD, MAX and MIN of Fls, and MAX of L, N, S and R; Component 3 mainly represents AVE indices of S and Ton. These results can also be seen in Figure 4.2.2, where the first three components are displayed. In Table 4.2.5, it also shows that when three components are retained from the 24 components, R STD and Ton MIN, which have relatively low proportion values of below 0.5, are not well represented. When four or five components are retained, the proportions of all the indices' variance that can be explained are high, which means that all these indices are well represented by the principal components.

Table 4.2.5 Component matrix and communalities for AVE, STD, MAX, and MIN indices of the six parameters

	Component Matrix					Communalities		
	Component 1	Component 2	Component 3	Component 4	Component 5	Extraction of 3 components	Extraction of 4 components	Extraction of 5 components
L AVE	**.793**	.519	-.155	.048	.015	.922	.924	.924
N AVE	**.839**	.477	-.121	.030	.083	.946	.947	.954
S AVE	.007	.520	**.659**	.339	.358	.705	.820	.948
Ton AVE	-.153	.526	**-.697**	.374	.010	.786	.926	.926
R AVE	**.865**	.362	.204	-.128	-.156	.920	.936	.960
Fls AVE	-.570	**.663**	.251	.181	-.304	.827	.860	.952
L STD	**-.665**	.385	-.071	-.415	.411	.595	.767	.936
N STD	.026	**.712**	-.296	-.455	.333	.595	.802	.913
S STD	**-.620**	.594	.259	.042	.239	.804	.806	.863
Ton STD	-.320	.542	-.575	.230	-.139	.726	.779	.798
R STD	-.234	.597	.077	-.590	-.229	**.417**	.765	.818
Fls STD	-.569	**.610**	.223	.008	-.291	.745	.745	.830
L MAX	.407	**.789**	-.203	-.202	.171	.829	.870	.899
L MIN	**.872**	.327	-.107	.151	-.119	.878	.901	.915
N MAX	.459	**.774**	-.172	-.253	.128	.840	.904	.920
N MIN	**.916**	.253	.010	.190	-.042	.902	.938	.940
S MAX	-.362	**.695**	.482	.156	.256	.846	.871	.936
S MIN	.536	.228	.577	.346	.227	.673	.792	.844
Ton MAX	-.449	.595	-.501	.275	-.053	.806	.881	.884
Ton MIN	.059	.211	-.542	.404	.156	**.342**	.505	.530
R MAX	.394	**.614**	.172	-.404	-.379	.563	.726	.869
R MIN	**.929**	.134	.230	.094	-.098	.933	.942	.952
Fls MAX	-.583	**.644**	.219	.070	-.238	.803	.808	.864
Fls MIN	-.340	**.602**	.215	.345	-.240	.524	.643	.701

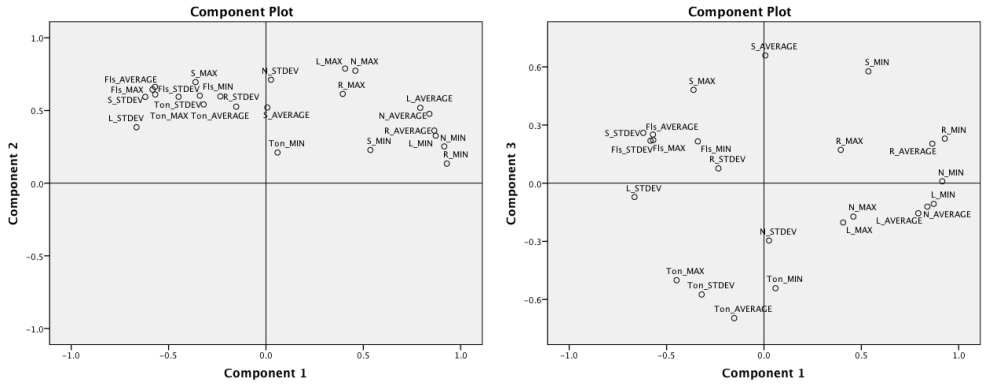


Figure 4.2.2 Loading plot of the principal components based on AVE, STD, MAX, and MIN data

4.2.2.3 Principal components analysis based on average and standard deviation indices of the five psychoacoustic parameters

Since SPL and loudness are highly correlated, e.g. L AVE and N AVE have a high correlation of greater than 0.9 as shown in Table 4.2.1; also since the information

contained in the data set of the MAX and MIN indices and that of the AVE and STD indices are similar, and the PCA based on AVE, STD, MAX and MIN data shows similar results to that based on AVE and STD data as discussed above, only AVE and STD data of the 5 psychoacoustic parameters are used in this section, without SPL data. Based on the AVE and STD data of the 5 parameters for the 102 recordings, the PCA is conducted on the correlation matrix of the 10 indices, from which 10 components are extracted. The sampling adequacy of KMO test is not high, above 0.5 as shown in Table 4.2.2. The first four components are retained to show the results here, while the variances (or eigenvalues) of the first three components are greater than 1. They respectively account for 38.1%, 22.7%, 17.0% and 9.5% of the total variance, shown in Table 4.2.6.

Table 4.2.7 shows the correlations between the first four principal components and the 10 indices. Component 1 has high correlations (above 0.6) with Fls AVE, S STD, Ton STD, R STD and Fls STD; Component 2 has high correlations with N AVE and R AVE; Component 3 has high correlations with S AVE and Ton AVE; Component 4 has high correlations with R STD. These results can also be seen on the component loading plots, as shown in Figure 4.2.3, where only the first three components are displayed. It can be seen that some of the indices are clustered in groups, e.g. Fls AVE, Fls STD and S STD; and N AVE and R AVE, where the indices in each group have high correlations among them, as shown in Table 4.2.1. It is clear that Component 1 mainly represents fluctuation related properties, including Fls AVE, Fls STD, and S STD; Component 2 mainly represents loudness related properties, including N AVE and R AVE; and Component 3 mainly represents sharpness and tonality properties.

Table 4.2.7 also shows the proportion of each index's variance that can be explained by the retained principal components. When 4 components are retained from the 10 components, the proportions of all the indices' variance that can be explained are high, which means that all these indices are well represented by the principal components. When 3 components are retained, all the indices have relatively high proportion values except for R STD, which has a value below 0.5. When 2 components are retained, S AVE, Ton AVE, Ton STD and R STD, which have relatively low proportion values of below 0.5, are not well represented. Overall, the first three components are generally necessary to represent the original 10 indices, which account for 77.5% of the total variance, while the first two components account for 60.8% and the first four components account for 87.0% of the total variance.

Table 4.2.6 Total variance explained by components based on AVE and STD indices of the five psychoacoustic parameters

Component	Eigenvalues	% of Variance	Cumulative %
1	**3.808**	**38.083**	**38.083**
2	**2.272**	**22.720**	**60.802**
3	**1.669**	**16.693**	**77.496**
4	.948	9.481	86.977
5	.593	5.934	92.911
6	.308	3.079	95.990
7	.168	1.676	97.666
8	.123	1.231	98.897
9	.077	.775	99.672
10	.033	.328	100.000

Table 4.2.7 Component matrix and communalities for AVE and STD indices of the five psychoacoustic parameters

	Component Matrix				Communalities		
	Component 1	Component 2	Component 3	Component 4	Extraction of 2 components	Extraction of 3 components	Extraction of 4 components
N AVE	-.088	**.956**	-.001	.184	.922	.922	.956
S AVE	.422	.222	**.638**	.507	**.228**	.635	.892
Ton AVE	.528	.226	-.710	.279	**.330**	.834	.912
R AVE	-.192	**.883**	.298	-.001	.816	.905	.905
Fls AVE	**.889**	-.165	.211	.093	.817	.861	.870
N STD	.540	.562	-.170	-.344	.608	.637	.755
S STD	**.841**	-.180	.270	.093	.740	.812	.821
Ton STD	**.640**	.090	**-.675**	.177	**.417**	.873	.905
R STD	**.622**	.233	.158	**-.641**	**.442**	**.467**	.878
Fls STD	**.850**	-.198	.204	-.041	.761	.803	.804

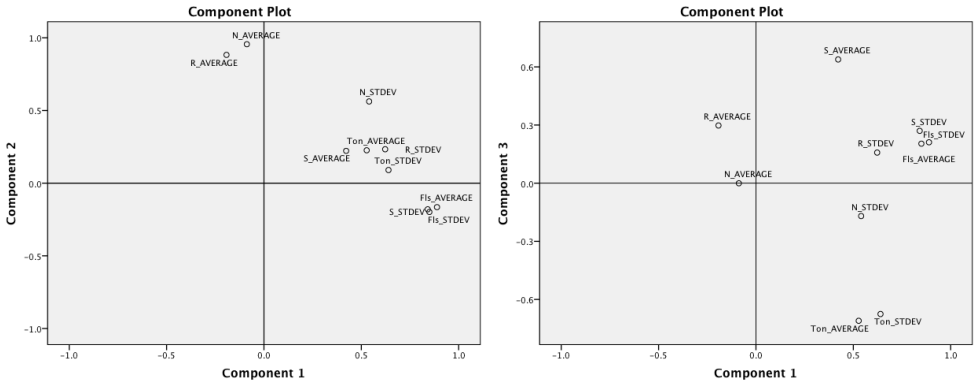


Figure 4.2.3 Loading plot of the principal components based on AVE and STD indices of the five psychoacoustic parameters

In sum, the PCAs based on the three different data sets generally show similar results, i.e., based on AVE and STD data and on all AVE, STD, MAX and MIN data,

with and without SPL data. It is understandable as the correlations between the MAX and MIN indices and the AVE and STD indices, and between loudness, roughness, and SPL. From the correlations between the principal components and indices described above, it is shown that generally Fls AVE, N AVE and Ton AVE/S AVE have the highest correlations with Components 1, 2 and 3 respectively, though the order of Components 1 and 2 may exchange. This suggests that the average values of fluctuation strength, loudness and tonality/sharpness (both of which are timbre features) are the key indices for characterizing sounds in soundscape, based on the present data set. This result corresponds to that of a previous research (De Coensel and Botteldooren 2006), which, on the basis of soundscape literature, suggests that three important physical indicators of soundscape are sound strength, spectral content, and temporal structure.

4.3 Characteristics of Sound Categories

To study the characteristics of each sound category, correlations between the sound categories and the principal components, as well as between the sound categories and the key indices, are examined, as shown in Table 4.3.1. Here, the first three principal components of the 10 indices in Section 4.2.2.3 are used, as well as Fls AVE, N AVE and Ton AVE/S AVE as key indices. It can be seen that, in terms of significant correlation at the 0.01 level, water sounds have negative correlation with fluctuation strength and tonality; wind sounds have negative correlations with fluctuation strength, tonality and sharpness; birdsongs have positive correlations with fluctuation strength and sharpness and negative correlation with loudness; and urban sounds have positive correlation with loudness and tonality. In this section, the results are based on the 101 sound samples, one less traffic sound from the 102 recordings (Yang and Kang 2010; 2013a). These results can also be visualized with more detailed information as presented below.

A three-dimensional coordinate system is established with its axes presenting the three principal components. Correspondingly, the 101 sound samples are visualized in Figure 4.3.1. The sound samples are also plotted in another coordinate system as shown in Figure 4.3.2, with the three key indices, the main indices that contribute to each principal component, as the axes, i.e. fluctuation strength AVE, loudness AVE and sharpness AVE (which is used here), although it should be noted there are some correlations between the three axes. From both figures it can be seen that the recordings in different categories generally centralize on different areas in the coordinate systems, although there are also overlaps. In other words, some clear differences among the different sound categories can be seen with the principal components and key indices.

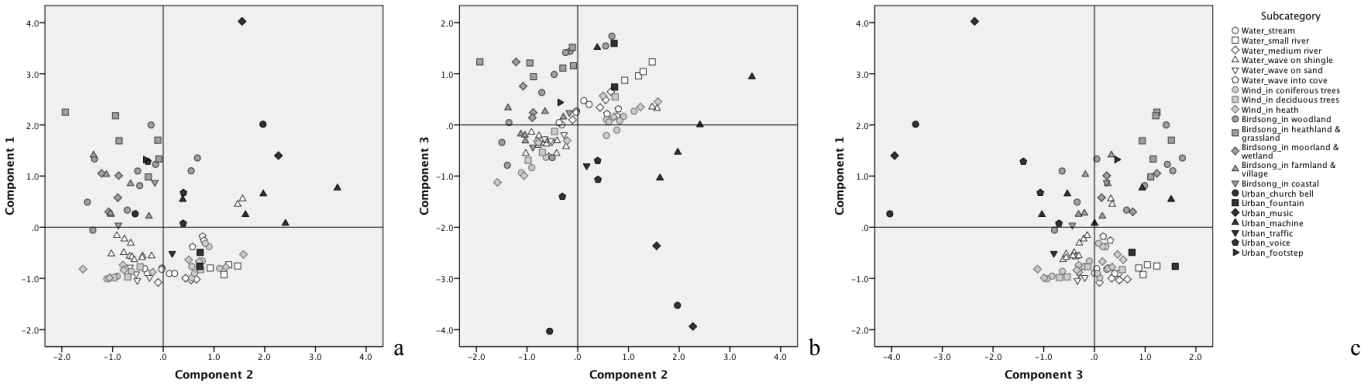


Figure 4.3.1 Component scores of 101 recordings in the three-dimensional coordinate system constituted by the first three principal components, (a) Components 1 and 2, (b) Components 2 and 3, (c) Components 1 and 3

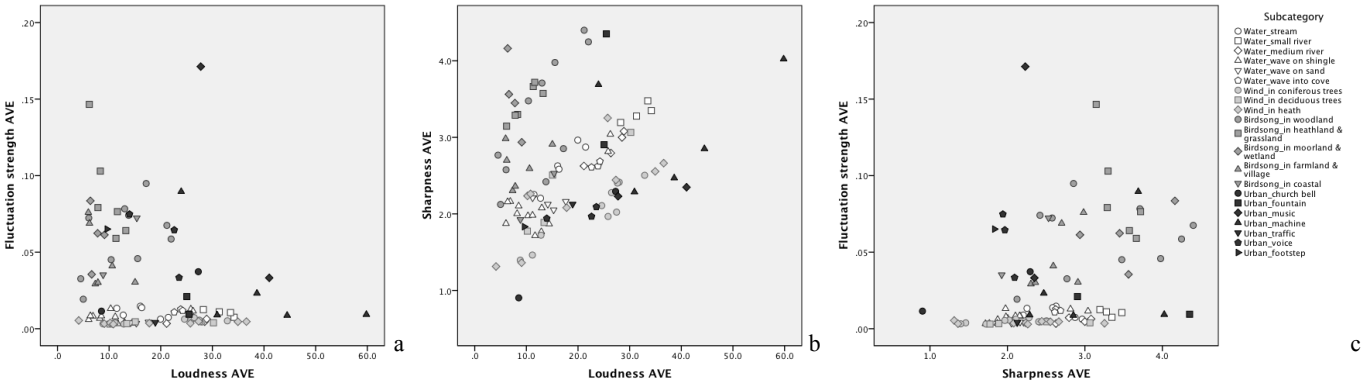


Figure 4.3.2 Component scores of 101 recordings in the three-dimensional coordinate system constituted by the three key indices, (a) Fls AVE and N AVE, (b) S AVE and N AVE, (c) Fls AVE and S AVE

Table 4.3.1 Pearson's correlation coefficients among the principal components, key indices, and sound categories, where ** indicates that correlation is significant at the 0.01 level (2-tailed), and * indicates that correlation is significant at the 0.05 level (2-tailed)

	Component 1	Component 2	Component 3	Fls AVE	N AVE	Ton AVE	S AVE	Water	Wind	Birdsong	Urban
Component 1	1.000										
Component 2	.000	1.000									
Component 3	.000	.000	1.000								
Fls AVE	.889**	-.165	.211*	1.000							
N AVE	-.088	.956**	-.001	-.199*	1.000						
Ton AVE	.528**	.226*	-.710**	.304**	.217*	1.000					
S AVE	.422**	.222*	.638**	.446**	.219*	-.019	1.000				
Water	-.442**	.079	.127	-.398**	.040	-.305**	-.095	1.000			
Wind	-.400**	.056	-.087	-.366**	.134	-.223*	-.367**	-.382**	1.000		
Birdsong	.634**	-.439**	.309**	.651**	-.434**	.093	.522**	-.435**	-.332**	1.000	
Urban	.249*	.363**	-.433**	.134	.319**	.524**	-.093	-.316**	-.241*	-.275**	1.000

In Figure 4.3.1 (a) (Components 1 and 2) and Figure 4.3.2 (a) (Fls AVE and N AVE), much of the areas of water sounds and wind sounds are overlapped, while birdsongs and urban sounds are generally located at separate areas, although some recordings categorized in urban sounds are mixed in the areas of birdsong, water and wind. In Figure 4.3.1 (a), in the urban sound category, one of voice (kids' voice in this case), footsteps and one of church bells recordings are in the birdsongs area; fountains and traffic are in the water and wind areas. Other human voices (adult) are located near the center of the coordinate system, while music, another church bells and most of machine sounds are at the top and right sides, far from the others. Two recordings of sea wave on shingle, which have the highest N AVE and R AVE in that subcategory, are in the area with most of urban sounds. In Figure 4.3.1 (b) (Components 2 and 3) and Figure 4.3.2 (b) (S AVE and N AVE), the four categories of sound are generally in different areas, although some parts overlap. With a given loudness value, sharpness of birdsong is higher than that of water sounds, and than that of wind sounds. In Figure 4.3.1 (b) fountains and footsteps in the urban sounds category are mixed with water sounds and birdsongs; one of the machine sounds is mixed with birdsongs; while the other urban sounds are separated from recordings in the three natural sounds categories.

Based on the results in Table 4.3.1 and Figure 4.3.1 and Figure 4.3.2, it can be seen that water sounds have low fluctuation strength average values and a wide range of loudness; wind sounds have low fluctuation strength average values, a wide range of loudness and low sharpness average values; birdsongs have high fluctuation strength average values, high sharpness average values and low loudness average values; and urban sounds have high loudness average values. Moreover, the results suggest that water

and wind sounds have similar psychoacoustic characteristics; fountains are similar to natural water sounds; traffic sounds are closer to water or wind sounds compared to other sounds. Music, church bells and machine sounds in the category of urban sounds are rather different from the natural sound groups.

In terms of the differences between natural and urban sounds, generally speaking, certain urban sounds are have fluctuation strength and loudness, while natural sounds have either low fluctuation strength and varied loudness and sharpness, or high fluctuation strength and sharpness and low loudness. Sounds with high fluctuation strength are birdsongs and those from human activity/facility, whereas sounds from non-life sources have low fluctuation strengths.

In the research of Hall et al. (2013), the relationship between acoustic and perceptual properties and listeners' judgements of sound quality has been investigated through multiple regression analyses. It was found that the regression models could not explained a large proportion of the variance of the two principle dimensions of the emotional and cognitive response to urban soundscapes, which were pleasantness and vibrancy, by the acoustical and psychoacoustical factors. However, in this study, results show that natural sounds, that are generally preferred by human, are not characterised by single dimension of any of the acoustical/psychoacoustical factors or combination of the factors. For example, natural sounds can either have low fluctuation strength and high loudness, or high fluctuation strength and low loudness. Thus, it provides a hint that sound preferences may be also not characterised by any single dimension of acoustical/psychoacoustical factors, but more complex pattern of them.

4.4 Categories Identification/Classification with Artificial Neural Networks and Discriminant Functions

While through the PCA some characteristics of different categories of sound are found, to automatically identify sound categories with the acoustical and psychoacoustical parameters, artificial neural network (ANN) models are explored. In addition, the results are compared with those by discriminant function analysis (DFA).

In this section, the 30-second segments of the recordings are used as cases for both ANNs and DFA. There are 1140 cases in total for the 101 recordings, among which, for neural networks, 150 are selected randomly for testing; with the remaining 990 cases used for training. A number of networks are designed based on different input data, and different layer structures are also considered.

4.4.1 Networks design and training

With all the six parameters, five networks are first developed based different input data and layer structures, among which one is based on AVE data of the parameters, two are based on AVE and STD data, and the other two are based on AVE, STD, MAX and MIN data. Table 4.4.2 shows the detailed information for the networks, including input data and network structures. For each network, the number of input nodes is equal to the number of input indices used in the model. Four output nodes are designed for each of the networks, as the recordings are to be identified into four categories: water, wind, bird and urban sounds. One or two hidden layers are chosen for the five networks. The number of nodes in each hidden layer has been optimized by adjusting the number and examining the nodes' percentage contributions to output signals of that layer, which show the degree of hidden nodes being utilized by the network, as shown in Table 4.4.1. If many nodes in a hidden layer of network contribute little or nothing to the output response, then that layer structure may be over designed with too many nodes; contrarily, if all hidden nodes show strong contributions, extra nodes may be needed to help the model. The network inputs are normalized automatically between 0 and 1 in order to improve training characteristics; the output, as in a binary form, does not require normalization. Full connection is chosen for all the networks, which means that all nodes in each layer receive connections from all nodes in each preceding layer. Sigmoid transfer function is selected for hidden and output layers in all the networks to normalize the nodes' output signal strength between 0 and 1.

Table 4.4.1 Percentage contributions of hidden nodes for Networks 1-5

Network Name		Network 1	Network 2	Network 3	Network 4	Network 5
Hidden Layer 1	Node 1	8.44	19.55	6.87	15.29	11.33
	Node 2	7.72	12.3	37.3	13.91	14.56
	Node 3	6.26	21.1	10.14	5.99	12.28
	Node 4	25.61	9.78	27.21	20.29	15.58
	Node 5	15.58	24.83	1.98	9.27	16.2
	Node 6	36.39	11.09	16.5	14.24	5.51
	Node 7	-	1.26	-	20.33	2.96
	Node 8	-	0.09	-	0.68	21.58
Hidden Layer 2	Node 1	-	-	13.86	-	13.47
	Node 2	-	-	2.62	-	15.85
	Node 3	-	-	17.75	-	1.38
	Node 4	-	-	0.31	-	1.33
	Node 5	-	-	16.34	-	18.93
	Node 6	-	-	15.99	-	14.7
	Node 7	-	-	12.37	-	16.17
	Node 8	-	-	20.75	-	18.17

Table 4.4.2 Network design and training information, and statistics results of Networks 1-8

Network				Network 1	Network 2	Network 3	Network 4	Network 5	Network 6	Network 7	Network 8
Input Data				AVE of 6 parameters	AVE, STD of 6 parameters	AVE, STD of 6 parameters	AVE, STD, MAX, MIN of 6 parameters	AVE, STD, MAX, MIN of 6 parameters	AVE, STD of N, S, Ton, R, and Fls	AVE, STD of S, Ton, R, and Fls	AVE, STD of S, Ton, and Fls
Network Architecture	Number of Layers			3	3	4	3	4	3	3	3
	Nodes of Input Layer			6	12	12	24	24	10	8	6
	Nodes of Hidden Layer 1			6	8	6	8	8	8	8	6
	Nodes of Hidden Layer 2			-	-	8	-	8	-	-	-
	Nodes of Output Layer			4	4	4	4	4	4	4	4
Training Information	Iterations			10000	10000	10000	10000	10000	10000	10000	10000
	Learn Rate			0.0046	0.0035	0.0021	0.0100	0.0100	0.0100	0.0100	0.0100
	Momentum Factor			0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
	Fast-Prop Coef			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Training Error			0.142	0.070	0.066	0.052	0.057	0.068	0.109	0.199
	Test Set Error			0.140	0.110	0.060	0.052	0.108	0.091	0.124	0.204
Correlation Statistics	Training	Correlation	Node 1	0.933	0.991	0.983	0.988	0.995	0.988	0.906	0.846
			Node 2	0.895	0.985	0.993	1.000	0.997	0.988	0.898	0.821
			Node 3	0.993	0.997	0.998	1.000	1.000	1.000	0.986	0.962
			Node 4	0.886	0.906	0.937	0.948	0.896	0.912	0.887	0.824
		Std Dev	Node 1	0.168	0.061	0.087	0.072	0.045	0.073	0.201	0.251
			Node 2	0.193	0.076	0.051	0.014	0.032	0.067	0.196	0.247
			Node 3	0.058	0.036	0.034	0.010	0.003	0.012	0.081	0.134
			Node 4	0.105	0.095	0.079	0.072	0.101	0.092	0.106	0.128
	Test	Correlation	Node 1	0.945	0.970	1.000	0.998	0.983	0.962	0.907	0.833
			Node 2	0.943	0.955	1.000	0.983	0.962	0.961	0.846	0.804
			Node 3	0.985	0.974	0.986	1.000	0.987	0.999	0.970	0.939
			Node 4	0.641	0.926	0.914	0.967	0.783	0.992	0.787	0.920
		Std Dev	Node 1	0.151	0.116	0.011	0.030	0.086	0.126	0.209	0.247
			Node 2	0.149	0.120	0.009	0.079	0.117	0.126	0.213	0.259
			Node 3	0.083	0.112	0.082	0.009	0.077	0.025	0.119	0.173
			Node 4	0.164	0.090	0.086	0.059	0.141	0.031	0.147	0.089

Table 4.4.3 Detailed prediction errors of Networks 1-8. For cases in Groups 1 and 2, ratios of the number of cases with high prediction error to the total number of cases in each of the four sound categories are displayed.

	Category	Predicted as	Network 1	Network 2	Network 3	Network 4	Network 5	Network 6	Network 7	Network 8
Group 1	Water	-	19/367 (94.8%)	3/367 (99.2%)	1/367 (99.7%)	0/367 (100.0%)	0/367 (100.0%)	4/367 (98.9%)	27/367 (92.6%)	63/367 (82.8%)
	Wind	-	49/283 (82.7%)	8/283 (97.2%)	1/283 (99.6%)	1/283 (99.6%)	2/283 (99.3%)	7/283 (97.5%)	52/283 (81.6%)	82/283 (71.0%)
	Birdsong	-	5/429 (98.8%)	1/429 (99.8%)	0/429 (100.0%)	0/429 (100.0%)	0/429 (100.0%)	0/429 (100.0%)	5/429 (98.8%)	12/429 (97.2%)
	Urban	-	16/61 (73.8%)	11/61 (82.0%)	7/61 (88.5%)	5/61 (91.8%)	13/61 (78.7%)	9/61 (85.2%)	20/61 (67.2%)	28/61 (54.1%)
Group 2	Water	Wind	4 wave on shingle, 2 wave into cove	1 wave on shingle	1 wave on shingle	-	-	2 wave on sand	1 wave on shingle, 7 wave on sand	1 medium river, 5 wave on shingle, 33 wave on sand, 2 wave into cove
		Birdsong	-	-	-	-	-	-	-	-
		Urban	-	-	-	-	-	-	-	-
		Total	6/367 98.4%	1/367 99.7%	1/367 99.7%	0/367 100.0%	0/367 100.0%	2/367 99.5%	8/367 97.8%	41/367 88.8%
	Wind	Water	12 wind in deciduous trees, 7 wind in coniferous trees, 12 wind in heath	1 wind in deciduous trees, 3 wind in heath	1 wind in heath	-	2 wind in heath	3 wind in coniferous trees, 1 wind in heath	14 wind in deciduous trees, 5 wind in coniferous trees, 9 wind in heath	5 wind in deciduous trees, 42 wind in coniferous trees, 5 wind in heath
		Birdsong	5 wind in heath	-	-	-	-	-	-	-
		Urban	-	-	-	1 wind in coniferous trees	-	-	-	-
		Total	36/283 87.3%	4/283 98.6%	1/283 99.6%	1/283 99.6%	2/283 99.3%	4/283 98.6%	28/283 90.1%	52/283 81.6%
	Birdsong	Water	-	-	-	-	-	-	-	1 birdsong in woodland, 1 birdsong in coastal
		Wind	-	-	-	-	-	-	-	-
		Urban	1 birdsong in woodland	-	-	-	-	-	-	1 birdsong in moorland and wetland, 1 birdsong in farmland and village
		Total	1/429 99.8%	0/429 100.0%	0/429 100.0%	0/429 100.0%	0/429 100.0%	0/429 100.0%	0/429 100.0%	4/429 99.1%
	Urban	Water	1 fountain, 2 footstep	5 fountain	5 fountain	5 fountain	1 fountain, 2 footstep	1 fountain	1 fountain, 2 traffic, 1 footstep	4 fountain
		Wind	4 fountain, 1 traffic	1 traffic	-	-	4 fountain, 1 machine, 1 footsteps	4 fountain	4 fountain	1 fountain, 3 traffic
		Birdsong	1 music, 3 machine, 2 footsteps	3 machine, 2 footsteps	1 machine, 1 voice	-	3 machine, 1 footsteps	3 machine	3 machine, 1 voice	5 music, 4 machine, 5 voice, 4 footsteps
		Total	14/61 77.0%	11/61 82.0%	7/61 88.5%	5/61 91.8%	13/61 78.7%	8/61 86.9%	12/61 80.3%	26/61 57.4%

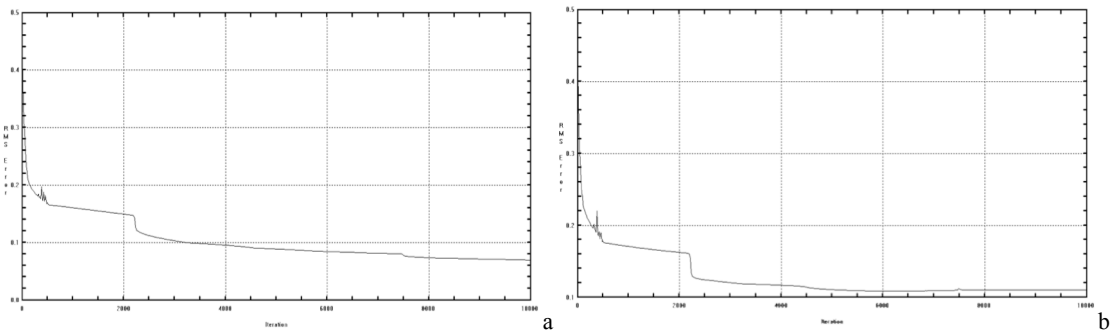


Figure 4.4.1 RMS error of (a) training set and (b) test set for Network 2

Part of the settings for training parameters is listed in Table 4.4.2. Complete histories of root-mean-square (RMS) error between the network's output response and the training targets monitored during training for both training and test set errors of the five networks show that they are generally at an appropriate level of training and are convergent. Figure 4.4.1 shows the histories of RMS error of Network 2 as an example. It is noted that as the errors are small for both training and test set, between 0.05 and 0.14 for the five networks, no additional effort are made to optimize the models, although further minor improvements are still possible by adjusting training parameters such as iterations, learning rate coefficient and momentum factor.

4.4.2 Network identifications of sound category

From the statistical comparison of network predictions with training targets, as shown in Table 4.4.2, it can be seen that for Output Nodes 1, 2 and 3, which are for identifying water, wind and birdsong categories respectively, the correlations for both training and test sets of the five networks are rather high, mostly over 0.90. For Output Node 4, for identifying urban sound category, the correlations vary between 0.64 and 0.96, which is lower than those for the other three output nodes, but still acceptable.

In order to examine the causes and details of the errors, the cases with relatively large prediction errors are analysed in Table 4.4.3. In the table, Group 1 includes cases that the difference between network prediction and target is over 0.4 in any of the four output nodes. Among these cases, Group 2 includes cases that the output node with the highest value among the four does not match the target node. Taking Network 2 as an example, cases in Group 2 show that one wave sound in the water category is classified as wind sound, four wind sounds are classified as water sound, while all the birdsongs are identified correctly. In the urban sound category, all fountains are classified into water

sound category, one traffic sound is classified as wind sound, and some footsteps and machine sounds are classified as birdsongs. These detailed errors may be expected, and they also approximately correspond to the results from PCA. On the other hand, the result that fountains, which are classified as urban sound in this paper, are identified as water sound, suggests that although created by human, fountains would not significantly differ from natural water sounds. In Group 1, the error cases of water, wind and bird sounds are below 3% except those predicted by Network 1, while the error cases of urban sounds vary from 8.2% to 26.2% for the five networks. The lower accuracy in identifying urban sound category is probably caused by the smaller sample size, and more importantly, the more complicated sound types.

In order to further explore the internal mechanism that the networks make the predictions, the key inputs, which are defined here if their contributions to an output node are above the average, are identified, including SPL AVE, N AVE, S AVE, R AVE, SPL MIN, R MIN for the output node in identifying water sounds; SPL AVE, N AVE, S AVE, R AVE, SPL MIN for identifying wind sounds; N AVE, S AVE, Fls AVE, Ton AVE, S STD, Fls STD, Ton MAX for identifying birdsongs; and SPL AVE, Ton AVE, Fls STD, Ton STD for identifying urban sounds. These indices correspond to the ones which demonstrate particular characteristics of each sound category obtained from PCA.

Among the five networks, Networks 2 and 3, which are both based on AVE and STD data as inputs but with different network structures, and Networks 4 and 5, which are based on AVE, STD, MAX and MIN data with different structures, have better performance than Network 1, which is based on AVE data only. The performance of Networks 2-5 does not differ greatly, although Networks 3 and 4 are slightly better than the other two. This, corresponding to the PCA results, suggests that the results based on AVE and STD data and those based on AVE, STD, MAX and MIN data are similar and thus, AVE and STD data are generally sufficient for the sound identification. Additionally, the number of hidden layers for network structure is not critical in this study. Overall, correlations between network predictions and targets using AVE and STD data as inputs are above 0.95 for the three natural sound categories and above 0.90 for the urban sound category, while the correlation of network using AVE data is above 0.90 for the three natural sound categories and above 0.64 for the urban sound category. The accurately predicted cases outside Group 1 (i.e. cases that the differences between network prediction and target are below 0.4), using AVE and STD data, are above 97% for the three natural sound categories and above 82% for the urban sound category, and by using AVE data only these figures are 83% and 74% respectively.

4.4.3 Networks based on part of the parameters

Additional networks have been developed based on part of the parameters. Network 6 is based on the AVE and STD of loudness, sharpness, roughness, fluctuation strength and tonality as inputs, as SPL and loudness are highly correlated for the current data set and, with and without SPL the PCA results remain almost the same as discussed before. The design information of the networks and prediction results are shown in Table 4.4.2. It can been seen that the correlations between network predictions and targets are high, all above 0.9 for the four categories of both training and test sets, similar to those of Networks 2 and 3, which suggest that the prediction ability is generally good.

While the SPL and loudness could depend strongly on the listening or recording positions, two more networks are designed, without SPL and loudness data included. Network 7 is based on the AVE and STD of sharpness, roughness, fluctuation strength and tonality as inputs. Since there is a relatively high correlation between roughness and SPL/loudness, Network 8 is constructed based only on the AVE and STD of sharpness, fluctuation strength and tonality. From Table 4.4.2, it can be seen that although the accuracy of these two networks are generally lower than the former six networks, they are still acceptable. Between these two, overall, Network 7 is slightly better than Network 8. For Network 7, the correlations are around 0.9 or higher for both training and test sets, except for Node 2 (0.85) and Node 4 (0.79) for the test set. For Network 8, the correlations are generally above 0.8. These two networks demonstrate that the sound categories can still be generally identified without loudness measures.

4.4.4 Discriminant function analysis

Compared to ANNs, DFA provides linear and simpler functions to identify the category of sound. Two sets of discriminant functions are presented here, respectively based on 10 indices, i.e., AVE and STD of loudness, sharpness, roughness, fluctuation strength and tonality, of the 1140 cases; and on 3 indices, i.e., loudness AVE, sharpness AVE and fluctuation AVE. For each condition, all input independents, or say the indices, are considered together, without stepwise method. Three functions have been developed respectively. The function coefficients for each variable of each function are shown in Table 4.4.4, as well as the group centroids for each function, all of which are used for the classification.

Table 4.4.4 Discriminant function coefficients and group centroids of functions based on psychoacoustic indices

			Function for 10 indices			Function for 3 indices		
			1	2	3	1	2	3
Discriminant Function Coefficients (Unstandardized)	Loudness AVE		-0.090	-0.066	-0.419	-0.123	0.086	0.054
	Sharpness AVE		1.082	0.390	1.431	1.513	-1.020	1.097
	Roughness AVE		0.000	0.821	3.987			
	Fluctuation AVE		8.761	-22.227	-28.404	18.760	33.152	-7.014
	Tonality AVE		4.320	16.951	22.517			
	Loudness STD		-0.191	0.107	0.489			
	Sharpness STD		3.637	-2.686	-2.664			
	Roughness STD		-0.113	1.308	-3.032			
	Fluctuation STD		-0.225	24.400	19.289			
	Tonality STD		11.150	25.707	-15.675			
	(Constant)		-2.936	-3.194	-5.058	-2.498	0.310	-3.605
Functions at Group Centroids	Category	Water	-1.373	-0.109	0.743	-0.888	-0.343	0.124
		Wind	-2.011	-0.418	-0.783	-1.724	0.056	-0.162
		Birdsong	2.598	-0.132	-0.075	2.076	0.038	-0.027
		Urban	-0.680	3.523	-0.312	-1.256	1.538	0.198

Table 4.4.5 Classification results by original discriminant functions based on psychoacoustic indices

	Category	Predicted Group Membership based on 10 indices				Predicted Group Membership based on 3 indices			
		Water	Wind	Birdsong	Urban	Water	Wind	Birdsong	Urban
Percentage %	Water	**77.4**	22.6	0.0	0.0	**87.7**	12.3	0.0	0.0
	Wind	8.8	**91.2**	0.0	0.0	20.5	**73.1**	0.0	6.4
	Birdsong	7.2	0.0	**91.8**	0.9	15.2	0.0	**83.2**	1.6
	Urban	36.1	13.1	4.9	**45.9**	8.2	27.9	13.1	**50.8**

From the predicted classification results shown in Table 4.4.5, it can be seen that the accuracies of the prediction are generally acceptable, although lower than the results by ANNs (Table 4.4.3). Comparing the prediction results by discriminant functions based on 10 indices and Network 6 in ANNs, for which the indices considered are the same, the numbers or proportions of correctly classified cases are less by the discriminant functions for all the four categories. The proportions by Network 6 are all around 99% for the three natural sound categories and 87% for urban sound category, while they are 77% to 92% and 46% by the discriminant functions. Between the two sets of discriminant functions, the proportions of correctly classified cases based on 3 indices are lower than those based on 10 indices for wind and birdsong categories, and higher for water and urban sound categories. The proportions of correctly classified urban sound cases by both sets are relatively small, about 46% and 51%. These results are tested through cross validations, in which each case is classified by the functions derived from all cases other than that one. The results in validations are almost the same as those by the original functions, which suggests the accuracies by the original functions presented above are reliable for the

current data set. Overall for the four categories, 84.6% of originally grouped cases and 84.5% of cross-validated grouped cases are correctly classified for the condition of 10 indices; and 80.4% and 80.2% of those for the condition of 3 indices.

4.5 Conclusions

Various natural and urban sounds, categorized as four main types, water, wind, birdsong, and urban, have been analysed with acoustical and psychoacoustic parameters, using HCA, PCA, ANN and DFA.

Three key indices, which are average values of fluctuation strength, loudness and sharpness, have been identified to show differences among different sound types. Water sounds have low fluctuation strength and a wide range of loudness; wind sounds have low fluctuation strength, a wide range of loudness and low sharpness; birdsongs have high fluctuation strength, high sharpness and low loudness; and urban sounds have high loudness. In terms of the differences between natural and urban sounds, generally speaking, urban sounds are with high fluctuation strength and loudness, while natural sounds are either with low fluctuation strength and varied loudness and sharpness, or with high fluctuation strength and sharpness and low loudness.

While the sound categories cannot be identified using any single acoustical and psychoacoustic parameter, identification can be made with a group of parameters. The ANNs have better performance than the discriminant functions for the classification. With ANNs, correlations between network predictions and targets using AVE and STD data as inputs are above 0.95 for the three natural sound categories and above 0.90 for the urban sound category. Without the influence of loudness, the correlations are still generally above 0.8 for the four sound categories.

Chapter 5
Applicability of pitch algorithms to environmental sounds

In this chapter, the applicability of pitch features and algorithms to environmental sound is explored. As expected, environmental sounds are mainly composed of complex tones and/or noises, rather than pure tones (sinusoids). The main purpose of this chapter is to look for a general model of pitch perception for all types of environmental sounds, not one model for one particular sound type.

A number of models are simplified/modified and implemented in Matlab program with MIRtoolbox, based on existing models of the two major classes of pitch theory as described in Chapter 2 Section 2.3.4, i.e., temporal models and spectral models, and of practical application in music and speech, discussed in Sections 5.2, 5.3 and 5.4, respectively. In Section 5.5, the performances of the implemented models for environmental sounds are compared, from which the one with best performances is selected for the further analysis of the current study. In Section 5.6, from the model selected, several descriptors (or parameters) are derived according to subjective pitch sensations. A number of basic statistic indices are then developed for the parameters to describe the variation of pitch characteristics of the sounds with time.

5.1 Introduction

5.1.1 Application of music features in soundscape

Over the past ten years, the perception and evaluation of soundscape have been investigated through numerous studies. It reveals that beside the A-weighted sound pressure level (SPL), additional parameters are necessary for soundscape measurement, e.g., background noise level, standard deviation of short L_{Aeq}, and psychoacoustic parameters, which are correlated to subjective evaluation like loudness, comfort, pleasantness, dynamics, and annoyance (Raimbault *et al.* 2003; Yang and Kang 2005a; Genuit and Fiebig 2006). Consequently, in this and the next chapters, in addition to the conventional parameters, more possible parameters are explored 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 (Schafer

1977), the applicability of music features – particularly the psychoacoustic parameters that have previously mainly been applied in music perception – in soundscape research are examined and demonstrated in this study with various types of common environmental sound (Yang and Kang 2011; 2013b). From alternative algorithms of the features proposed in literature for both perception of the auditory system and practical application in music and speech, the algorithms applicable for environmental sound are searched through comparison. With a number of parameters regarding these music features derived from respective corresponding algorithms selected, the different characteristics of various environmental sounds are explored in Chapter 7, as well as the parameters' contribution to the automatic identification of environmental sound type.

In the field of psychology of music, relations between music features and humans' emotion and evaluation have been studied for decades as reviewed in Chapter 2 Section 2.4.3. For example, fast tempo, rapid changes in loudness, sharp amplitude envelope with rapid attack and decay, loud music, high pitch, and wide pitch range may be associated with emotions like high activation, excitement, happiness, tension, anger, and fear; slow tempo, few or no changes in loudness, round envelope, soft music, low pitch, narrow pitch range, may be associated with low activation, calmness, sadness, boredom, pleasantness, and fear. Large pitch variation and simple, consonant harmony may be associated with happiness, pleasantness, activity, or surprise; small pitch variation and complex, dissonant harmony with sadness, unpleasantness, tension, anger, fear, or boredom (Hevner 1937; Watson 1942; Krumhansl 1996; 1997; Balkwill and Thompson 1999; Gabrielsson and Lindström 2001). The apparent contradiction may depend on the context, that is, the combination and interaction with other structural factors (Gabrielsson and Lindström 2001). It is expected that these parameters would benefit further study of soundscape evaluation, especially the emotional responses (Schulte-Fortkamp *et al.* 2007).

5.1.2 Music features

As reviewed in Chapter 2, the sensations of hearing are generally studied from four aspects in psychoacoustics and psychology of music, i.e., loudness, pitch, rhythm and timbre, among which the former three respectively correspond to physical aspects of sound of amplitude, frequency and time, while timbre corresponds to both frequency and time aspects.

While the characteristics of different types of environmental sounds have been analysed in terms of loudness and timbre aspects in the Chapter 4, including psychoacoustic parameters of loudness, sharpness, tonality, roughness, and fluctuation

strength, applicability of parameters in the remaining aspects, i.e., pitch and rhythm, to soundscape research are studied in this chapter and Chapter 6, respectively. In this study, the pitch and rhythm features are termed as music features to be distinguished from the previous psychoacoustic parameters. Consequently, in Chapter 7, the differences among various environmental sounds are studied in these music features.

5.1.3 Methods

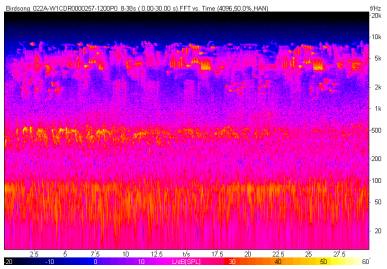
As the pitch and rhythm algorithms were developed for perception of the auditory system or application in music and speech, their applicability to soundscape study is examined with 11 environmental sounds with single sound sources and 2 soundscape sounds with mixed sound sources, representing typical natural and urban sounds. The 11 environmental sound recordings are sounds of stream, river, sea waves, wind, birdsong, fountain, church bells, street music, street machines, traffic, and voice. The 2 soundscape recordings are soundscape on a street with clock and traffic, and that in a park with fountain and geese. These 13 recordings were all made with a mono channel, a duration of 30 seconds, and a sample rate of 44,100 Hz (16 bit). The spectra over time of the recordings are shown in Table 5.2.1, from which it can be seen that the music has rather different spectrum with the other sounds especially, suggesting the potential inapplicability of music features to environmental sounds.

In this study, first, based on a systematic review of algorithms of pitch and rhythm proposed in literature for both perception of the auditory system and application in music and speech, a number of models are simplified/modified and implemented in Matlab program with MIRtoolbox. Then, the parameters of algorithms are adjusted for environmental sounds by balancing the results of the 13 sounds. Third, the performances of these implemented models for the environmental sounds are then compared, from which the ones with the best performances are selected for the further analysis in this study. Finally, a number of descriptors (or parameters) and statistic indices are derived based on the models selected.

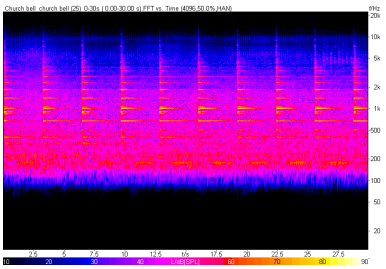
5.2 Temporal Models

Pitch models according to temporal theories of pitch perception are first considered, which are a class of pitch theories that based upon the extraction of timing information from auditory nerve activity (see Chapter 2 Section 2.3.4).

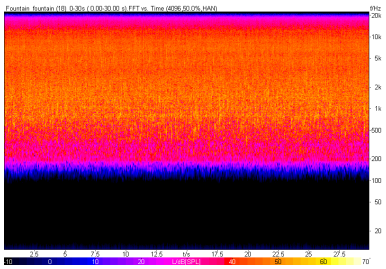
Table 5.2.1 Spectra over time of the 13 recordings



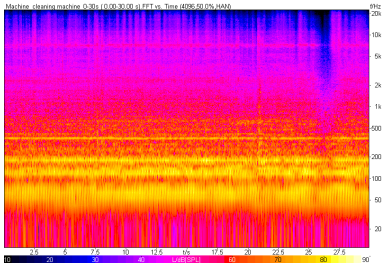
Birdsong_022A-W1CDR0000257-1200P0_8-38s



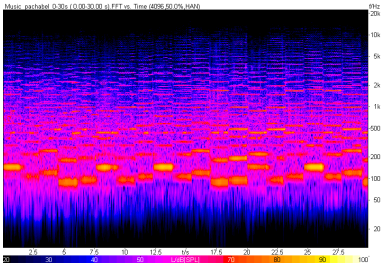
Church bell_church bell (25)_0-30s



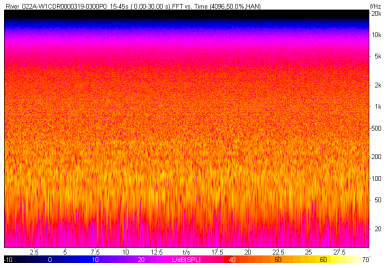
Fountain_fountain (18)_0-30s



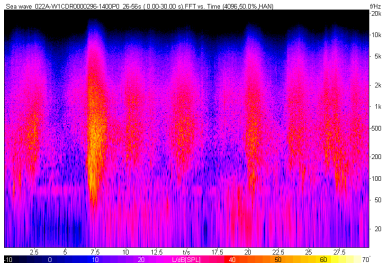
Machine_cleaning machine_0-30s



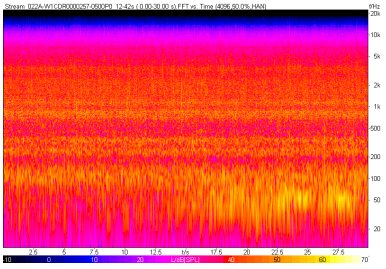
Music_pachabel_0-30s



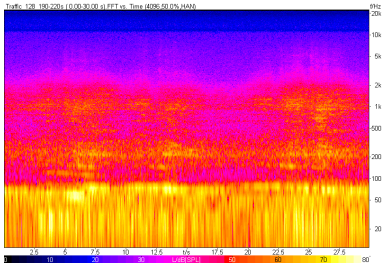
River_022A-W1CDR0000319-0300P0_15-45s



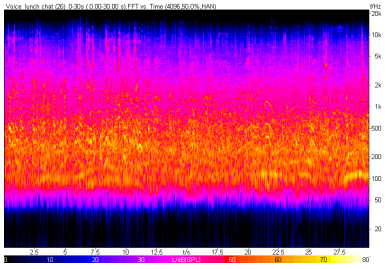
Sea wave_022A-W1CDR0000296-1400P0_26-56s



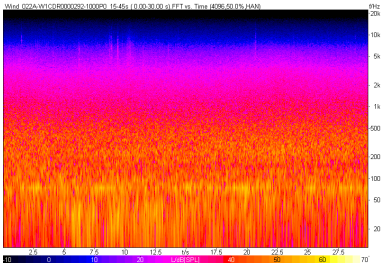
Stream_022A-W1CDR0000257-0500P0_12-42s



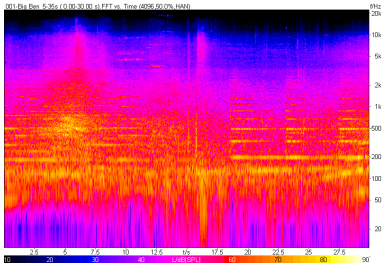
Traffic_128_190-220s



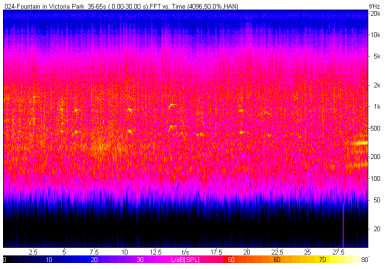
Voice_lunch chat (26)_0-30s



Wind_022A-W1CDR0000292-1000P0_15-45s



001-Big Ben_5-35s



024-Fountain in Victoria Park_1:00am_35-65s

5.2.1 Temporal models in literature

Based on the temporal patterns of firing in auditory nerves, Moore (1977) has proposed a model that incorporates features of the former temporal models and of the pattern cognition models, and thus, explains two pitch perception mechanisms associated with both resolved and unresolved harmonics, and account for the majority of experimental results in pitch perception (Moore 1997) (see Section 2.1.4). In his model (Moore 1997, p211-212), it is assumed that a complex stimulus first passes through a bank of filters (critical bands), the outputs of which produce activity in neurones with corresponding characteristic frequencies (CFs). In each channel, the temporal pattern of activity, i.e. the time intervals between successive nerve impulses, is analysed in a limited range that varies from about 0.5/CF to 15/CF seconds, which is the appropriate range for the time intervals that would occur most often (Moore 1997, p205). Then it searches across channels for common time intervals. The most prominent time interval among them is selected, and the perceived pitch corresponds to the reciprocal of the final interval selected (Moore 1997, p211-212).

Before Moore, Licklider (1951) has proposed a duplex theory of pitch perception which is somehow similar. He suggested that the auditory system employs both frequency analysis and autocorrelational analysis. The frequency analysis is performed by the cochlea or basilar membrane (BM) of acoustics stimulus. The neural part of the system performs the autocorrelational analysis of the trains of nerve impulses, which results in an autocorrelation function for each channel of the cochlear output.

Based on these pitch theories, Meddis et al. (Meddis and Hewitt 1991a; 1991b; Meddis and O'Mard 1997) have presented a quantitative model, which is a simplification and computational implementation of these theories or models. This model firstly simulates the band-pass filtering effect of the BM in the inner ear with critical-band filters, the same as that in Moore's model (Moore 1977; 1997). The output of each filter, which corresponds to the mechanical motion of the BM at that point, is then converted to a probability of spike occurrence in auditory nerve. This nonlinear mechanical to neural transduction at the hair cell is achieved approximately by a compressed half-wave rectification and low-pass filtering of the filtered input. The model computes an autocorrelation analysis, similar to that proposed by Licklider (1951), on the firing probabilities in each channel, to estimate the distribution of intervals between spikes in groups of auditory nerve fibres. All these autocorrelation functions (ACFs) of nerve fibre firing probabilities are summed across channels to generate a summary autocorrelation function (SACF), from which the pitch is predicted (Assmann and Summerfield 1990).

Full details of the model (Meddis and Hewitt 1991a; 1991b; Meddis and O'Mard

1997) consists of a number of stages: (1) The combined effects of outer and middle ear on frequency is simplified and implemented using a digital band-pass filter with skirts down by 3 dB at 450 and 8500 Hz. (2) The mechanical frequency selectivity of the cochlea is achieved using a set of 60 or 128 digital critical-band (gammatone) filters (Patterson *et al.* 1988). The centre frequencies of the overlapping filters are equally spaced (approximately 0.25 equivalent rectangular bandwidths (ERBs) apart for 128 gammatone filters), along a log scale between about 100 and 8000 Hz. (3) The probability of spike occurrence within the population of fibres of a bank of hair cells is approximately a compressed, half-wave rectified version of the filtered input for channels carrying a high-amplitude filter output; for low-amplitude filter outputs, however, the output follows the input with little obvious rectification or compression. This characteristic reflects the nonlinear transduction of mechanical motion of the BM to neural response in nerve fibre. Moreover, as the probability of spike generation in a fibre depends on recent history of firing of that fibre, an adjustment to the firing probability as a function of the time since it last generated a spike is computed. For medium intensity, continuous stimuli, however, the refractory effects of firing of auditory nerve fibres make very little difference to the probability of firing. (4) A distribution of time intervals among all spikes – not just successive spikes – is estimated within each channel, since this is a kind of approximation and is computationally convenient. The interspike interval histogram is computed in the form of calculation of a running autocorrelation function (ACF) of the nerve fibre firing probabilities over a short summation time (7.5 ms was used in the model). (5) Channel ACFs are summed or averaged across channels to generate a summary ACF (SACF). (6) The pitch is determined either by inspection of major peaks of the SACF – the perceived pitch corresponds to the highest peak, or by pitch matching, i.e. estimation of discriminability of stimuli, which is implemented by computing distance measures between of the SACFs of the stimuli.

5.2.2 Implementation of temporal models

Based on the model of Meddis et al. (Meddis and Hewitt 1991a; 1991b; Meddis and O'Mard 1997), a simplified model can be implemented with MIRtoolbox in Matlab. First, signal is decomposed through critical-band filter banks. In each channel, an autocorrelation function (ACF) is computed based on the filterband waveform. Then a summary autocorrelation function (SACF) is obtained by averaging the ACFs across channels or filterbands. Finally, the peaks of the SACF are picked, the abscissas (lag time of the function) of which correspond to the reciprocal of the values of the pitches of the signal, while the ordinates (autocorrelation coefficient) are related to the corresponding

pitch strengths of the pitches (see Section 5.3). In this simplified model, the ACFs are calculated directly based on the waveforms of the channels, without half-wave rectification or compression processing of the waveforms, since these processes are not available in MIRtoolbox at this stage. It is expected, however, this simplification would not effect systematically the modelling result. The procedure of this model is illustrated in Figure 5.2.1, using the stream sound in the 13 recordings. The figure (a), (b), (c) and (d) respond to the four steps of the model respectively, i.e. filterbank decomposition, ACFs of each of the filterbands, the SACF, and peak selection. In figure (d), the peaks selected are indicated by small circles. For illustration, ten gammatone filters with half overlapping along a scale between 50 and 22000 Hz are used here. The ACFs and thus pitches are computed over the whole duration of the signal, i.e. 30s. The procedure is, however, the same for the case of short summation time, when signal is first decomposed into successive frames of short duration, and pitches are calculated within each frame.

With this method, the pitch/pitches of the 13 sound samples are calculated, based on both the whole duration and successive frames of short duration of the signals. Part of the results based on successive frames is displayed in Table 5.5.1, which shows the variation of pitches over time. The frame length of 46.4 ms and hop length of 10 ms are used according to Tolonen and Karjalainen (2000). Table 5.5.2 shows the average values of pitches and the corresponding pitch strengths of the sound samples over the whole duration of 30s. In both conditions, the best four pitches are shown.

A number of auditory filters are available, including the gammatone filterbank, filterbank along the Bark scale and an approximation of critical bands with third-octave band filters, to simulate the response of the basilar membrane (BM). The gammatone filters distribute linearly along a frequency scale measured in equivalent rectangular bandwidths (ERBs), which means that the width of each band is in proportion to the filter centre frequency. The shape of the gammatone filter (fourth order) is very similar to that of the roex filter, which is commonly used to represent the magnitude characteristic of the human auditory filter (Patterson and Moore 1986; Patterson *et al.* 1988; 1992). This filterbank in matlab calls the Auditory Toolbox routines MakeERBFilters and ERBfilterbank (Slaney 1993; Lartillot 2011). Here, 10, 20, 40 and 80 gammatone filters are used respectively to compare the performances. The Bark scale filterbank here uses the band edges of 100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400, 7700, 9500, 12000 and 15500 Hz. The third-octave band filters, as an approximation of critical bands, have been used in loudness calculating procedure (ISO 532 1975; Zwicker and Fastl 1999). The third-octave band filterbank used here consists of 21 non-overlapping bands which covers the frequencies from 44 Hz to 18 kHz. Since for lower frequencies (below about 300 Hz), third-octave

bands are too small in relation to the critical bands, three third-octave bands are added together to approximate a critical band. That is, the lowest three filters among the 21 are one-octave band-pass filters, while the remaining eighteen are third-octave band-pass filters. The band edges respond to 44*[2^([0:2, (9+(0:17))/3])] according to Klapuri (1999). The filterbanks are implemented using elliptic filters with order of 4. The algorithms based on these different filterbanks are compared in Section 5.5 to explore the one of optimal performance.

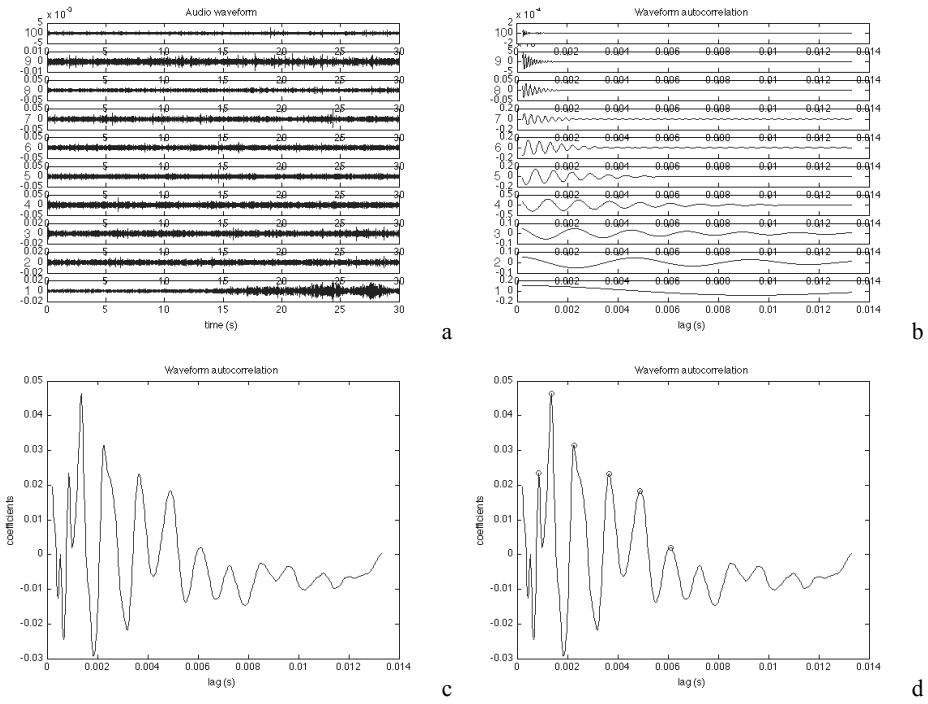


Figure 5.2.1 The procedure of implemented temporal pitch model, illustrated with a stream sound. (a), (b), (c) and (d) correspond to the commands as below respectively in Matlab:

f=mirfilterbank('Stream','Gammatone','Hop',2)
ac=mirautocor(f,'Min',0.0002,'max',0.0133)
s=mirsum(ac,'Mean')
p=mirpeaks(s,'Threshold',0.4,'Contrast',0.1,'NoBegin')

For ACF computation, normal autocorrelation is used, i.e., without nonlinear compression of amplitude in the frequency domain (also see Sections 5.3 and 5.4). It is equal to that when the parameter of exponential compression is set to a value of 2 (Tolonen and Karjalainen 2000). While it is desired to specify the range of period of pitch (lag time of ACF) taken into consideration in each channel separately according to Moore (1997, p205), which is limited and varies with the centre frequency of the channel, it has difficulty to be done with MIRtoolbox. Thus, the pitch range considered in each channel

is unitary, between 75 and 5000 Hz. The maximum pitch frequency is set to 5000 Hz, because above 4-5 kHz the ability to discriminate changes in the frequency of pure tones diminishes and the sense of musical pitch disappears, and also the tones produced by musical instruments, the human voice and most everyday sound sources all have fundamental frequencies below this range (Moore 1997).

For specifically selecting the peaks of the SACF, a number of conditions are made and different values of the condition parameters have been compared based on the results of the 13 sounds. Before peak picking, the total amplitude of the SACF is normalized between 0 and 1, corresponding to the minimum and maximum of the SACF. In other words, a distance of 1 is equivalent to the distance between the maximum and the minimum of the function (Lartillot 2011). It is postulated that a given local maximum will be considered as a peak if its normalized amplitude is higher than a threshold, specified by the parameter of "*Threshold*". A second condition is that a given local maximum will be considered as a peak if the differences of amplitude with respect to both the previous and successive local minima are higher than a threshold, specified by the parameter of "*Contrast*". In addition, the first sample is ignored, not considered as a possible peak candidate, since all stimuli produce maxima near the place where lag time is equal to zero, which is a general property of autocorrelation functions (Wightman 1973). Since the peaks should not be considered as real peaks if their autocorrelation values are negative, taking the algorithm based on 10 gammatone filters for example, the value of the parameter of "Threshold" is set to 0.4, balancing the results of the 13 sounds that the picks of which the absolute amplitude are below 0 are generally removed. The parameter of "Contrast" of 0.1 is used based the results of the 13 sounds. That is, only the maxima of which normalized amplitude is higher than 0.4 and the differences of amplitude with both the previous and successive local minima are higher than 0.1 are selected. The data used for parameters setting of the algorithms based on the various filterbanks is shown in Appendix I, from which the values of the parameters determined are indicated with corresponding commands in Section 5.5. While multiple pitches can be picked with no limitation of number of peaks, the best four pitches are extracted, corresponding to the highest four peaks. The number of pitches is generally enough to show the pitch properties of the sounds.

5.3 Spectral Models

Pitch models based on spectral (place) theories that are available are considered additionally. They assume that as the spectral analysis performed in the inner ear, the

frequencies of a sound are represented by the excitation of different places (neurones) along the BM, and the pitch is determined by the pattern of excitation (see Chapter 2 Section 2.3.4).

5.3.1 Spectral models in literature

Following the principles of virtual pitch theory (Terhardt 1974b), Terhardt et al. (1982) have proposed an algorithm to extract pitch from sounds, and also the pitch strength. The algorithm is based on both the spectral-pitch pattern and the virtual-pitch pattern. Each of these patterns consists of pitch (height) values and associated pitch weights, which account for the relative pitch strength of every individual pitch. The whole pitch percept is described as a combination or competition of the two pitch patterns. The spectral-pitch pattern is constructed by spectral analysis, extraction of tonal components, evaluation of masking effects, and weighting according to the principle of spectral dominance (Ritsma 1967). This specific algorithm is described in Chapter 2 Section 2.3.7, in which the tonality is calculated based on spectral-pitch pattern. The virtual-pitch pattern is deduced from the spectral-pitch pattern by a process of subharmonic coincidence assessment (Terhardt *et al.* 1982). For each spectral pitch present, the subharmonics are calculated. Virtual pitch is deduced by a scanning mechanism for the coincidence (or near coincidence) of all the subharmonics components. Each virtual-pitch components is evaluated by a weight which takes into account the number of subharmonics components coincident, the weight of the relevant spectral-pitch components, the accuracy of the near coincidences, and the existence region of virtual pitch (fundamental frequency) (Ritsma 1962). Then the spectral-pitch pattern and the virtual-pitch pattern are combined by approximately evaluating the competition between spectral and virtual pitches in the way that the original spectral-pitch weights are reduced by a factor of about 0.5.

While Terhardt's pattern recognition model is computationally sophisticated, a model which is relatively easy for implementation can be found in the theory of Wightman (1973), which shows a family similarity to those of Terhardt (1974b; 1979) and Goldstein (Goldstein 1973; de Boer 1977). The "pattern-transformation model" proposed by Wightman (1973) is a mathematical model based on pattern recognition of pitch perception. It is a spectrally based autocorrelation model (frequency domain computation), and thus is phase-insensitive, which is different from temporally based autocorrelation models (time domain computation) that are phase-sensitive (Licklider 1951) (described in Section 5.2). The autocorrelation is computed by Fourier transform of the power spectrum of the temporal waveform. First, an acoustic stimulus is transformed

into a pattern of peripheral neural activity, which roughly represents the power spectrum of the stimulus. Then a Fourier transformation is performed on the peripheral pattern, thus the output pattern roughly represents the autocorrelation function of the stimulus. Pitch is derived from the positions of maximal activity in the transformed pattern, i.e. the positions of the highest peaks in the autocorrelation function, except those near the place corresponding to the abscissa of zero of the autocorrelation function. The strength of a pitch is thought to be related to the absolute height of the corresponding peak in the pattern (Wightman 1973).

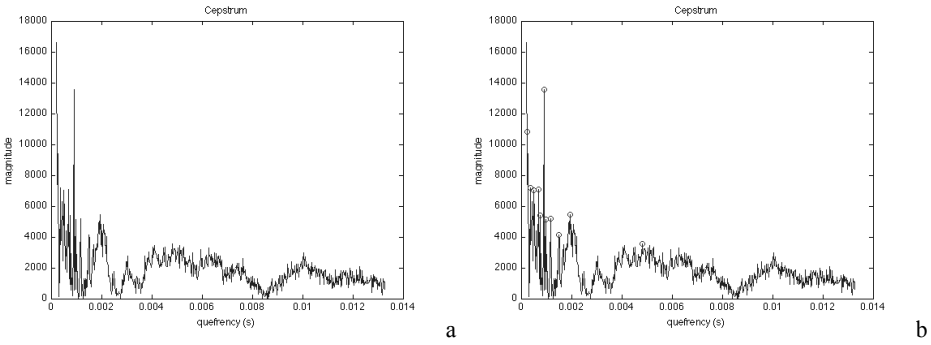
5.3.2 Implementation of spectral models

For the computational simplicity, Wightman's pitch model is implemented here with MIRtoolbox in Matlab. One implementation is by calculating first the spectrum of a signal and then the spectrum of the spectrum. Since spectrum of a spectrum of signal is equal to a cepstrum, an alternative way of implementation is through the calculation of cepstrum of a signal. *Cepstrum*, termed by Tukey (Bogert *et al.* 1963), is defined as the power spectrum of the amplitude-logarithm of the power spectrum, and has been used for pitch detection in voiced-speech (Noll 1964; 1967). Figure 5.3.1 shows an example of the algorithm based on cepstrum using a same stream sound as that in Figure 5.2.1. First, the cepstrum of sound is calculated, as shown in Figure 5.3.1 (a) of the stream sound for the whole duration of 30s, in which the lowest and highest values in quefrency time been calculated correspond to the pitch range of 75 to 5000 Hz. The pitches respond to the peaks in the cepstrum, shown in Figure 5.3.1 (b).

For peak picking, in addition to the two conditions used in the temporal model in Section 5.2, a third condition is applied. It postulates that only peaks with abscissa distance greater than a given threshold are considered. The threshold, specified by the parameter of "*Reso*", is set to semi-tone, i.e. the ratio between the two peak positions is equal to 2^(1/12). The higher peak remains out of two conflicting peaks. Thus, it removes the peaks whose abscissa distance to one or several higher peaks is lower than a semi-tone (Lartillot 2011). The parameter of "Threshold" is set to 0. The values of the parameter of "Contrast" have been compared based on the results of the 13 sounds; a value of 0.2 is used. Similar to the temporal model, the first sample is ignored, not considered as a possible peak candidate, and the best four pitches are extracted, corresponding to the highest four peaks.

The pitch/pitches of the 13 sound samples calculated with this method are shown in Table 5.5.1 and Table 5.5.2. Table 5.5.1 shows the variation of pitches over time, the

results based on successive frames of short duration. The same frame length and hop length are used as in the temporal model. Table 5.5.2 shows the average values of pitches and the corresponding pitch strengths of the sound samples over the whole duration of 30s.



a b

Figure 5.3.1 The procedure of implemented spectral pitch model based on cepstrum, illustrated with a stream sound. (a) and (d) correspond to the commands as below respectively in Matlab:

c=mircepstrum('Stream','Min',0.0002,'max',0.0133)
p=mirpeaks(c,'Threshold',0,'Contrast',0.2,'Reso','SemiTone','NoBegin')

5.4 Simplification Models

Some simplification models for pitch detection have been developed for the application in music and speech. They were developed not for simulation of auditory perception but for practical application of real-time pitch analysis, and thus are computationally efficient and may be applicable for large sample size analysis.

5.4.1 Simplification pitch models in literature for music and speech

Tolonen and Karjalainen (2000) have proposed a pitch analysis model which can be seen as a computationally simplification of the model of Meddis and O'Mard (1997). Instead of multi-channels in Meddis and O'Mard's model, this model essentially divides the signal into two channels, below and above 1000 Hz. It is thus more computationally efficient, and thought to has very similar behaviour to a certain extent for complex tones.

In the model, a pre-whitening filter is first used to remove short-time correlation of signal, using warped linear prediction model. It corresponds to flattening the spectral envelope in the spectral domain, which may be seen to have minor resemblance to

spectral compression in the auditory nerve. Then the signal is separated into two channels, below and above 1 kHz. The high-channel signal is half-wave rectified and lowpass filtered with a similar filter to that used for separating the low channel. The model computes a "generalized" autocorrelation of the low-channel signal and of the envelope of the high-channel signal. The generalized autocorrelation, somehow similar to Wightman's approach, consists of the computation of a discrete Fourier transform, magnitude compression of the spectral representation, and an inverse transform. It thus allows the use of nonlinear processing and the control of the parameter of the frequency domain compression, e.g., the application of logarithm results in the cepstrum, which is not directly possible with time domain method. While for normal autocorrelation function the parameter of exponential compression is equal to 2, a value of 0.67 is used in the model to improve the detection performance (Tolonen and Karjalainen 2000).

The two autocorrelation functions, of the low-channel and of the high-channel, are summed to produce a summary autocorrelation function (SACF). To be more selective, the SACF is further processed to obtain an enhanced summary autocorrelation function (ESACF), which removes repetitive peaks with double the time lag – also multiples the time lag of factors of three, four, five, etc. – where the basic peak is higher than the duplicate, and also the near-zero time lag part of the SACF curve. The peaks in the SACF and ESACF curves indicate the pitches of signal (Tolonen and Karjalainen 2000).

5.4.2 Modification and implementation of simplification models

MIRtoolbox in Matlab has implemented part of Tolonen and Karjalainen's model with pre-specified command, without providing the pre-whitening filter. The parameter of compression for generalized autocorrelaion is set to 0.67 as recommended by Tolonen and Karjalainen. It computes the autocorrelation in the frequency domain and includes a magnitude compression of the spectral representation. A value of compression lower than 2 decreases the width of the peaks in the autocorrelation curve, at the risk however of increasing the sensitivity to noise (Tolonen and Karjalainen 2000). Thus, for environmental sounds in this study, a value of 2 is used instead, which corresponds to a normal autocorrelation. Also, the ESACF in Tolonen and Karjalainen's model is not used here; pitch information is instead indicated by the peaks in the SACF curve.

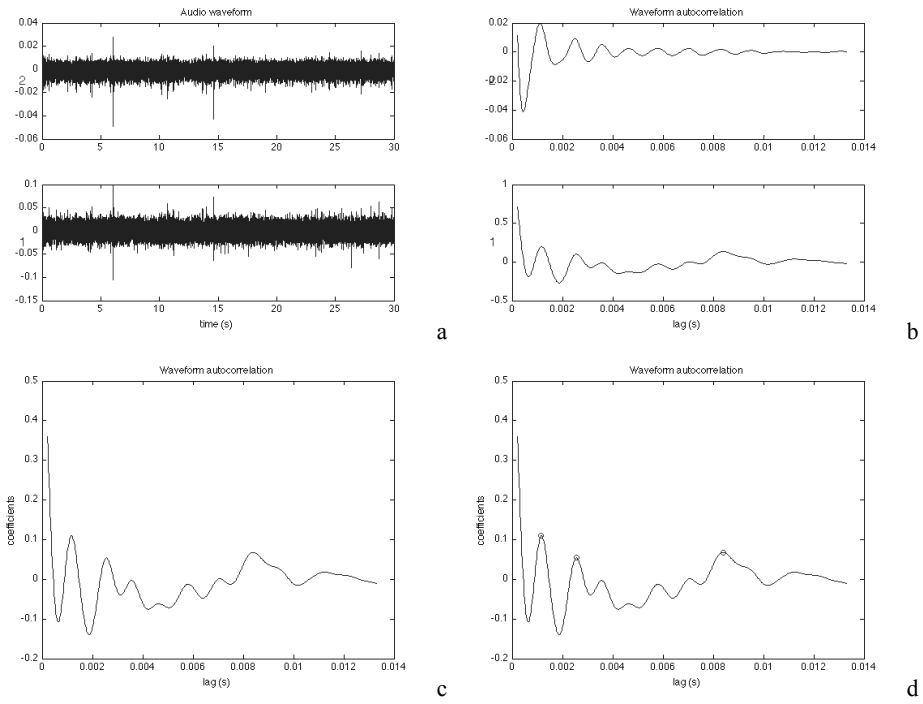


Figure 5.4.1 Procedure of the modified pitch model of Tolonen and Karjalainen, illustrated with a stream sound. (a), (b), (c) and (d) correspond to the commands as below respectively in Matlab:

f=mirfilterbank('Stream','2Channels')
ac=mirautocor(f,'Min',0.0002,'max',0.0133)
s=mirsum(ac,'Mean')
p=mirpeaks(s,'Threshold',0.3,'Contrast',0.1,'NoBegin')

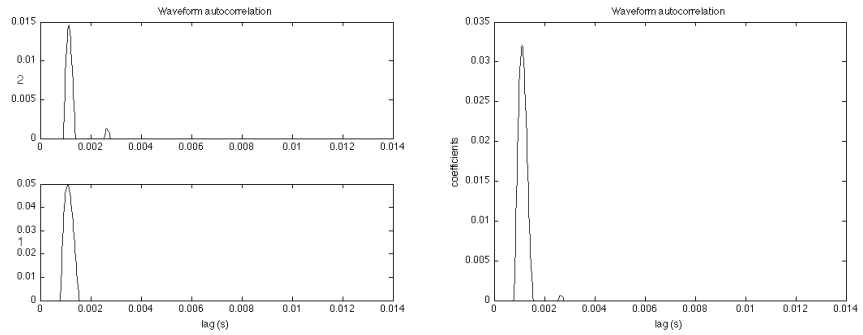


Figure 5.4.2 The procedure of pitch model according to Tolonen and Karjalainen, illustrated with a stream sound, corresponding to the commands as below respectively in Matlab:

ac=mirautocor(f,'Min',0.0002,'max',0.0133,'Enhanced', 2:10, 'Generalized', 0.67)
s=mirsum(ac,'Mean')

The procedure of the modified model is shown in Figure 5.4.1, using the same stream sound as in Figure 5.2.1 and Figure 5.3.1. First, the signal is decomposed into two

channels, below and above 1 kHz. The high-channel signal is lowpass filtered at cut-off of 1 kHz. Then a normal autocorrelation is computed based on the signal in each channel, in the pitch range of 75 to 5000 Hz, the same as that in two models above. The resulting two ACFs are averaged to produce a SACF. Pitches of signal respond to the peaks in the SACF, which are selected according to a number of conditions. The figure (a), (b), (c) and (d) respectively respond to these four steps. In the figure, the ACFs and thus pitches are computed over the whole duration of the signal (i.e. 30s).

Figure 5.4.2 shows the second and third steps of the model according to the pre-specified command in MIRtoolbox based on Tolonen and Karjalainen's model. The figures correspond to (b) and (c) in Figure 5.4.1. The parameters of the computation of ACF are set to 0.67 for compression and 2:10 for enhancement of the ESACF. A value of 2:10 for the parameter of "Enhanced" responds to that the original autocorrelation function is half-wave rectified, timescaled by factors from 2 to 10, and subtracted from the original clipped function (Lartillot 2011). It can be seen from the figures that only the peak that corresponds to high pitch is retained.

The values of the parameters of the conditions for peak picking have been compared based on the results of the 13 sounds. The values of 0.1 for "Contrast" and 0.25 for "Threshold" are used. Similarly, the first sample is ignored, not considered as a possible peak candidate, and the best four pitches are extracted, corresponding to the highest four peaks.

The pitch/pitches of the 13 sound samples calculated with this method are shown in Table 5.5.1 and Table 5.5.2. Table 5.5.1 shows the variation of pitches over time, the results based on successive frames of short duration. The same frame length and hop length are used as in the temporal and spectral models. Table 5.5.2 shows the average values of pitches and the corresponding pitch strengths of the sound samples over the whole duration of 30s.

5.5 Model Comparison

In order to compare the performance of the models implemented, the pitch/pitches of the 13 sound samples are calculated with these three types of models. The algorithms of these models are expressed respectively with single command with MIRtoolbox in Matlab, instead of the commands step by step, although the results have slight differences which can be ignored. The pitch results and the corresponding commands are shown in Table 5.5.1 and Table 5.5.2, as well as the parameters setting of each algorithm indicated in each command. It is noted that a number of the commands are not directly available in

MIRtoolbox, since the program has been modified to a small extent by the author to meet the needs in this study. Table 5.5.1 shows the results of the variation of pitches over time, based on calculation of successive frames, with three different models, one in each type. The frame length of 46.4 ms and hop length of 10 ms are used according to Tolonen and Karjalainen (2000). In each frame, the best four pitches (if they exist) are shown, i.e. the ones with the highest pitch strengths. The first pitch, the one with the highest pitch strength, is indicated by blue symbol, while the second pitch, i.e. the pitch with the second highest pitch strength, is indicated by green symbol, and similarly the third by red and the fourth by cyan. Table 5.5.2 shows the average values of pitches over the whole duration of 30s, using all the models, and the corresponding pitch strengths of the sound samples. The pitches are calculated by ACFs based on the whole duration, contrast with successive frames of short duration, and the best four pitches are shown. From Table 5.5.1 and Table 5.5.2, it can be seen that the results are quite different based on the different algorithms, though some matches.

5.5.1 Comparison of pitch models

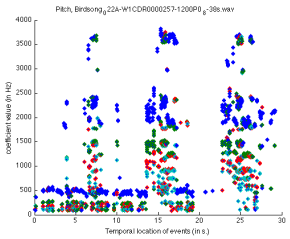
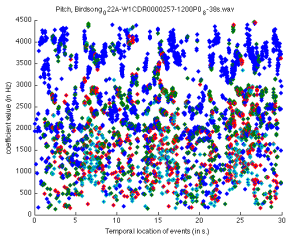
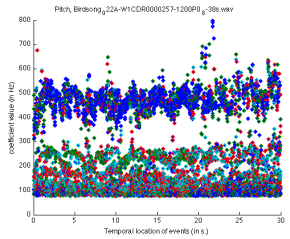
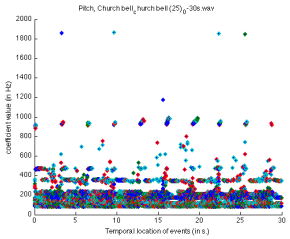
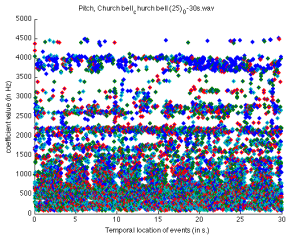
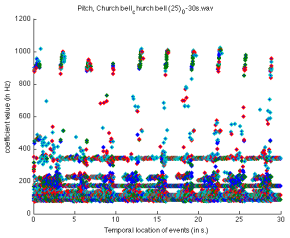
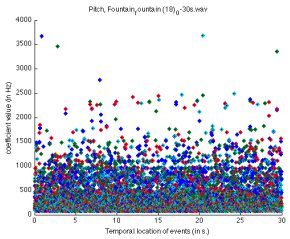
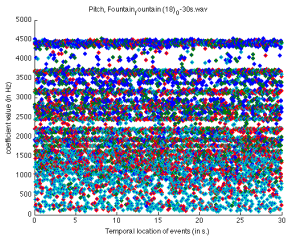
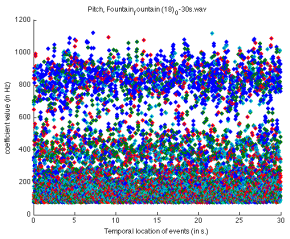
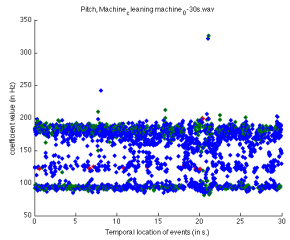
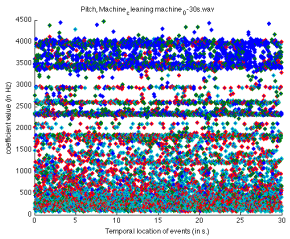
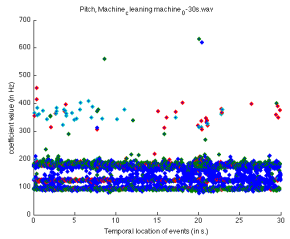
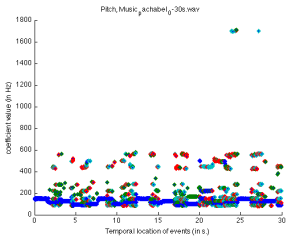
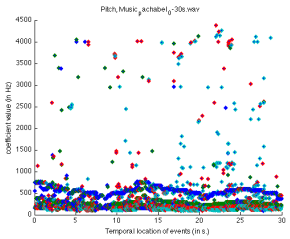
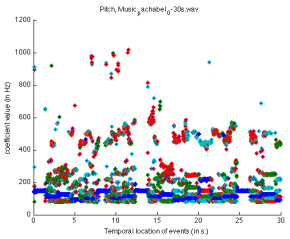
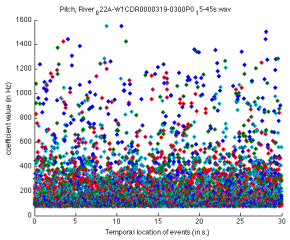
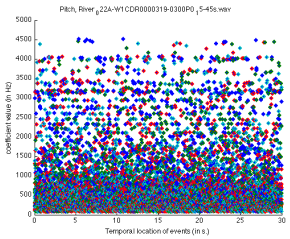
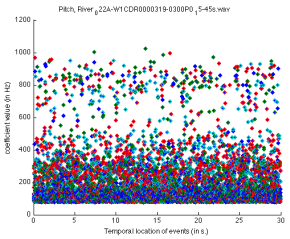
The simplification model of Tolonen and Karjalainen, i.e. the '2Channels' method, has the limited analysis range of pitch frequency of the maximum of about 1 kHz. In other words, this pitch analysis method focuses on low-to-mid fundamental frequencies and cannot derive pitch frequencies above about 1 kHz (Tolonen and Karjalainen 2000). It is determined by the boundary of the two channels, because both channels have the low-pass characteristics at the frequency of 1 kHz. It can be seen from the tables, using this method, that all the 13 sounds have the results of pitch values of around or below 1 kHz. However, different from most music and speech, the pitches of environmental sounds may exceed that region as expected, e.g., the birdsongs may have pitches much higher than 1 kHz, reach about 4 kHz with the temporal methods. Since the temporal method is the basic on which the two channels method is a simplification, and thus may derive more accurate results. It is clear that for environmental sounds, especially such as birdsongs, other method is needed. The two channels method, therefore, may not be used for pitch analysis of environmental sounds because of its limitation.

With the spectral model based on computation of cepstrum, it can be seen from the results shown in the tables that the pitch values of most of the 13 sounds are high. In Table 5.5.1, which shows the variation of pitches over time, the first pitches, or say the most prominent pitch, over a large amount of time reach 3 to 4 kHz for all the sounds except for music (around 500 Hz). In terms of average pitches over the whole duration,

shown in Table 5.5.2, the first pitches show the values of above 2.5 kHz for most of the sounds; the values of the first pitches are around 1 kHz for sea wave, stream and voice, and about 500 Hz for music. A possible reason of these relatively high pitch results is that environmental sounds may consist of large amount of noise, rather than pure or complex tones. The algorithm of Wightman's (1973) model, however, though can generate predictions about pitch of any signal, focuses on the analysis of complex tones. For noises, the predictions of high pitch values may result from the quick changes in spectra along frequency scale, while the power spectra of complex tones consist of evenly spaced components. The random change of noise signal produces high correlations at short time in autocorrelation function, which are interpreted as pitches with this algorithm but may not correspond to real pitch sensation. Although these inadequacies of the model may be corrected with a number of modifications in algorithm, e.g., in Tolonen and Karjalainen's (2000) model, which also involves computation of generalized autocorrelation in separate channels, a pre-whitening filter is used to remove short-time correlation of the signal, these modifications are either not available or too complicated (Wightman 1973). Thus, the current cepstrum method can only be used for pitch analysis of some sound types like music and speech, but not for general environmental sounds.

Since both the simplification model and spectral model may not be applicable to environmental sounds as discussed above, the temporal models may be the only option. Indeed, temporal models proved to be capable of explaining the majority of experimental results in pitch perception (Moore 1977), including both complex of tonal components and interrupted noise (Meddis and Hewitt 1991a). It can be seen from Table 5.5.2 that the values of the pitches calculated by the temporal models vary among the 13 sounds, from about 100 to 4000 Hz or no pitch perceived. Taking the results by the algorithm based on 40 gammatone filter bands for example, the values of the first pitches are about 4000 Hz for birdsongs, around 1000 Hz for sounds of fountain and stream, and about 100 to 200 Hz for the sounds left, except for river, sea waves, traffic and wind, in which no pitch is perceived. In terms of pitch strength, the first pitches of birdsongs and church bells have relative strengths of above 7.5, while that of music is about 4.9. They are about 1.5 to 2.3 for stream sound, soundscapes at Big Ben, and in Victoria Park, and below 0.9 for sounds of fountain, machine and voice. Somehow, these results may correspond to what could be expected such that the pitch value of birdsongs is high, river, sea waves, traffic and, wind may not produce any pitch, and pitch strengths of birdsongs, church bells and music are higher than the others.

Table 5.5.1 Pitches over time of 13 sounds using three different types of method

	Klapuri *p=mirpitch('folder','Klapuri','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0,'frame','total',4)*	Cepstrum *p=mirpitch('folder','Cepstrum','Max',5000,'Threshold',0,'Contrast',0.2,'Reso','SemiTone','frame','total',4)*	2Channels *p=mirpitch('folder','2Channels','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Generalized',2,'Enhanced',0,'frame','total',4)*
Birdsong			
Church bells			
Fountain			
Machine			
Music			
River			

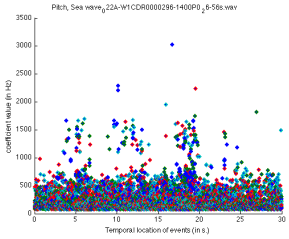
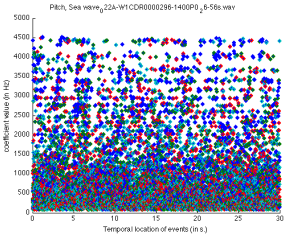
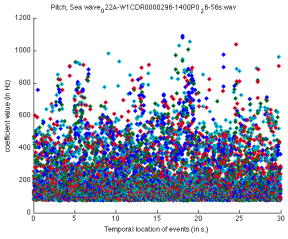
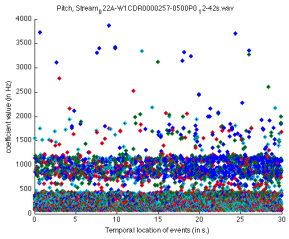
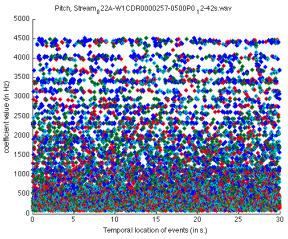
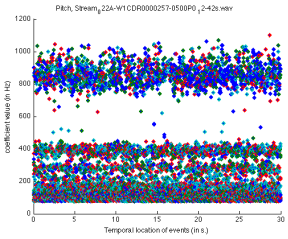
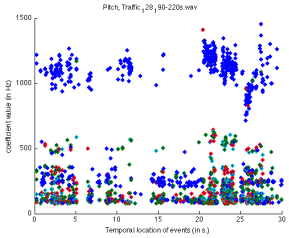
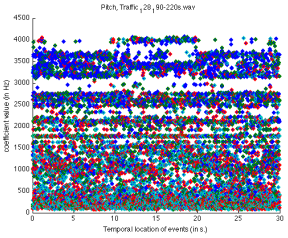
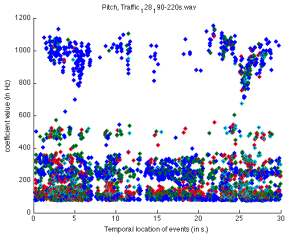
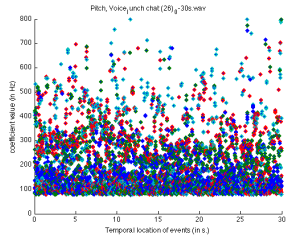
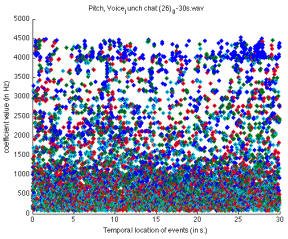
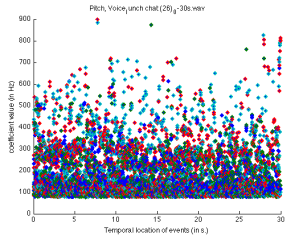
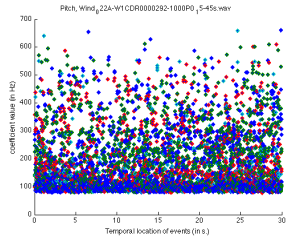
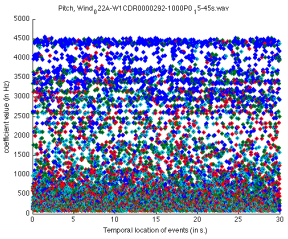
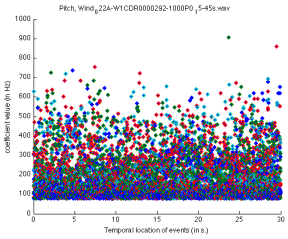
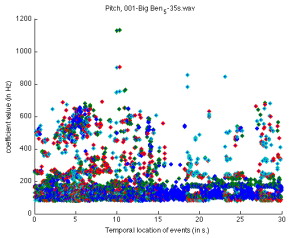
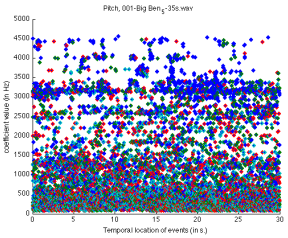
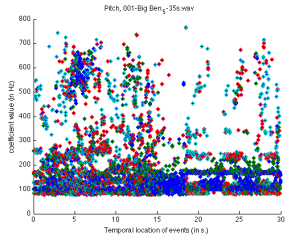
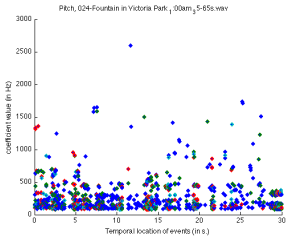
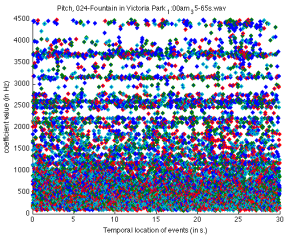
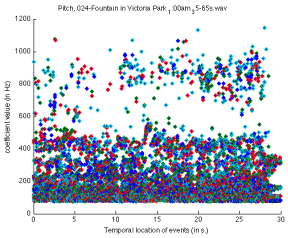
	Klapuri	Cepstrum	2Channels
Sea waves			
Stream			
Traffic			
Voice			
Wind			
Big Ben			
Victoria Park			

Table 5.5.2 Average pitches and corresponding pitch strengths of 13 sounds with different methods

		Birdsong	Church bells	Fountain	Machine	Music	River	Seawaves	Stream	Traffic	Voice	Wind	Big Ben	Victoria Park
Cepstrum	Pitch *p=mirpitch('folder','Cepstrum','Max',5000,'Threshold',0,'Contrast',0.2,'Reso','SemiTone','total',4)* *pv=mirgetdata(p)*	3815 1991 1095 -	2610 2158 3990 512	4421 3658 2754 3148	2329 3907 2575 1825	557 2486 1430 491	3163 2316 801 605	1002 1189 1078 2741	1105 3990 2742 1427	3574 2700 1770 1522	882 1463 4344 2094	4437 3650 - -	3142 2604 1288 1001	2567 3682 1154 2141
	Pitch strength (1.0e+04) *pa=get(p,'Amplitude')* *for i=1:13* *pa{1,i}{1,1}{1,1}* *end*	8.781 4.590 2.310 -	2.777 2.776 2.380 2.243	6.875 6.205 5.547 5.540	2.672 2.744 1.487 1.107	2.769 1.770 1.707 1.679	8.574 5.359 4.567 3.738	7.095 6.974 6.942 6.020	1.363 1.082 0.722 0.713	2.906 2.593 1.711 1.430	1.615 1.577 1.509 1.395	2.734 1.870 - -	2.362 1.591 1.429 1.246	2.087 1.766 1.562 1.443
2Channels	Pitch *p=mirpitch('folder','2Channels','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Generalized',2,'Enhanced',0,'total',4)* *pv=mirgetdata(p)*	483 85 248 102	114 101 129 939	823 266 394 -	174 94 - -	146 - - -	- - - -	- - - -	861 119 394 -	- - - -	272 - - -	- - - -	99 168 - -	77 133 91 470
	Pitch strength *pa=get(p,'Amplitude')* *for i=1:13* *pa{1,i}{1,1}{1,1}* *end*	1.689 0.644 0.571 0.522	2.096 1.502 1.454 1.433	0.847 0.233 0.225 -	0.551 0.420 - -	1.275 - - -	- - - -	- - - -	0.689 0.403 0.280 -	- - - -	0.230 - - -	- - - -	0.589 0.426 - -	0.361 0.358 0.276 0.269
10 Gammatone 'hop' 2	Pitch *p=mirpitch('folder','Gammatone','Max',5000,'Threshold',0.4,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0,'total',4)* *pv=mirgetdata(p)*	3672 1917 457 533	232 464 101 177	733 451 1159 276	200 93 - -	221 148 111 446	205 732 445 1189	445 206 112 720	735 442 1161 274	- - - -	218 444 111 154	214 439 - -	216 99 449 -	449 721 226 114
	Pitch strength *pa=get(p,'Amplitude')* *for i=1:13* *pa{1,i}{1,1}{1,1}* *end*	1.914 1.706 1.316 1.084	1.942 1.835 1.769 1.615	1.008 0.620 0.601 0.390	1.185 0.622 - -	1.634 1.082 1.044 1.036	0.758 0.748 0.747 0.627	1.149 0.807 0.336 0.285	1.407 0.980 0.763 0.707	- - - -	1.403 0.691 0.581 0.243	1.209 0.762 - -	1.532 0.792 0.147 -	1.366 0.758 0.679 0.411
20 Gammatone 'hop' 2	Pitch *p=mirpitch('folder','Gammatone','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0, 'total',4)* *pv=mirgetdata(p)*	4017 2011 453 491	205 102 93 473	739 360 459 274	104 - - -	147 111 558 209	105 153 357 275	153 206 275 355	923 745 276 365	- - - -	152 103 279 -	- - - -	102 153 - -	153 354 272 215
	Pitch strength *pa=get(p,'Amplitude')* *for i=1:13* *pa{1,i}{1,1}{1,1}* *end*	1.581 0.832 0.796 0.641	1.830 1.685 1.576 1.491	0.258 0.253 0.245 0.234	0.226 - - -	0.955 0.603 0.269 -0.097	0.253 0.180 0.111 0.072	0.443 0.315 0.261 0.254	0.554 0.427 0.351 0.322	- - - -	0.695 0.501 0.226 -	- - - -	0.607 0.341 - -	0.566 0.479 0.474 0.352
40 Gammatone 'hop' 2	Pitch *p=mirpitch('folder','Gammatone','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0, 'total',4)* *pv=mirgetdata(p)*	4072 2027 493 556	115 102 938 84	813 204 258 134	176 93 - -	148 - - -	- - - -	- - - -	1064 402 118 286	- - - -	276 - - -	- - - -	99 169 - -	153 77 134 91
	Pitch strength *pa=get(p,'Amplitude')* *for i=1:13* *pa{1,i}{1,1}{1,1}* *end*	8.618 4.714 4.616 3.702	7.529 7.130 7.027 6.728	0.634 0.384 0.302 0.288	0.770 0.675 - -	4.897 - - -	- - - -	- - - -	1.550 1.151 0.509 0.330	- - - -	0.857 - - -	- - - -	2.268 1.393 - -	1.516 1.476 1.410 1.201
80 Gammatone 'hop' 2	Pitch *p=mirpitch('folder','Gammatone','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0, 'total',4)* *pv=mirgetdata(p)*	4050 2017 491 551	115 102 939 84	812 205 268 376	177 93 - -	148 - - -	- - - -	- - - -	1067 400 119 281	- - - -	274 - - -	- - - -	99 169 552 -	77 153 134 91
	Pitch strength *pa=get(p,'Amplitude')* *for i=1:13* *pa{1,i}{1,1}{1,1}* *end*	275 158 146 122	3652 3511 3449 3294	37 27 22 19	17 14 - -	3629 - - -	- - - -	- - - -	77 53 46 9	- - - -	649 - - -	- - - -	270 195 146 -	738 697 636 618

		Birdsong	Church bells	Fountain	Machine	Music	River	Seawaves	Stream	Traffic	Voice	Wind	Big Ben	Victoria Park
Bark	Pitch *p=mirpitch('folder','Bark','Max',5000,'Threshold',0.3,'Contrast',0.15,'Reso','SemiTone','Compress',2,'Enhanced',0,'total',4)* *pv=mirgetdata(p)*	4078	102	745	125	149	130	-	1067	1122	275	-	100	134
		490	115	381	-	-	-	-	117	248	-	-	169	152
		552	949	101	-	-	-	-	398	117	-	-	-	77
		2036	84	278	-	-	-	-	278	-	-	-	-	90
	Pitch strength *pa=get(p,'Amplitude')* *for i=1:13* *pa{1,i}{1,1}{1,1}* *end*	2.217	1.470	0.149	0.716	0.894	0.170	-	0.205	0.508	0.230	-	0.674	0.359
		1.309	1.457	0.085	-	-	-	-	0.190	0.354	-	-	0.537	0.346
		1.118	1.344	0.083	-	-	-	-	0.190	0.192	-	-	-	0.333
		1.077	1.283	0.077	-	-	-	-	0.069	-	-	-	-	0.313
Klapuri	Pitch *p=mirpitch('folder','Klapuri','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0,'total',4)* *pv=mirgetdata(p)*	4085	102	751	177	149	-	94	887	-	94	-	99	151
		2035	115	374	93	-	-	-	402	-	276	-	169	134
		490	84	472	-	-	-	-	118	-	-	-	-	469
		549	941	949	-	-	-	-	297	-	-	-	-	77
	Pitch strength *pa=get(p,'Amplitude')* *for i=1:13* *pa{1,i}{1,1}{1,1}* *end*	1.922	1.841	0.170	0.311	1.313	-	0.170	0.377	-	0.232	-	0.472	0.262
		1.072	1.735	0.161	0.293	-	-	-	0.256	-	0.175	-	0.276	0.260
		0.969	1.682	0.131	-	-	-	-	0.175	-	-	-	-	0.234
		0.887	1.607	0.107	-	-	-	-	0.093	-	-	-	-	0.192

5.5.2 Comparison of filterbanks

Though the temporal models may be the appropriate method for pitch analysis of environmental sounds, there might be the problem that they are computationally expensive when a large number of filter bands are used. It is expected that the larger the number of filters used the more accurate the result would be – e.g. 60 or more gammatone filters have been used in Meddis et al.'s model (Meddis and Hewitt 1991a; 1991b; Meddis and O'Mard 1997). However, computation time increases with increasing number of filters. Hence, the performances based on different numbers of gammatone filters are compared, in order to look for a balance between computational efficiency and accuracy. Here, comparisons are made among 10, 20, 40 and 80 gammatone filters, the results of average pitches based on which are shown in Table 5.5.2. It shows that the results somehow differ when different numbers of filters are used. In terms of pitch values, 40 and 80 gammatone filters generally produce similar results; the pitches computed by both of these filters differ from and are more dispersive along the frequency scale than those by 20 or 10 gammatone filters. For example, the pitches that varied over time calculated based on 10 and 20 gammatone filters congregate at certain frequencies for sounds of fountain, river, sea waves, stream and wind, probably caused by the very limited number of filterbands. It can be concluded from the results that 40 gammatone filters are generally an optimization which produce relatively accurate results without too large number of banks.

Furthermore, additional types of auditory filters are available to be compared with the gammatone filters, which are the Bark scale filters and third-octave band filters. For the 13 sounds, the results of average pitches based on the two filterbanks are shown in

Table 5.5.2. It can be seen that the values of the average pitches calculated based on both the Bark scale and the third-octave band filters are similar to those based on 40 and 80 gammatone filters, though the order of the four most prominent pitches – i.e. sorted by pitch strength from high to low – may exchange. It can also be seen that the results calculated by the temporal methods with either the 40 and 80 gammatone, Bark scale or third-octave band filters are somewhat similar to those by the '2Channels' method, except for high frequencies above about 1kHz that cannot be simulated by the '2Channels' method. This agreement between the two types of models from another aspect supports the reliability of the pitch results. In terms of pitch strength, the relative strengths of the 13 sounds are also similar among the 40 gammatone, Bark scale and third-octave band filters, and also '2Channels' method; all these results may correspond to what could be expected of human's perception of pitch in environment. Among the filterbanks, both of the results by the Bark scale and the third-octave band filters are more approximative to those by 40 and 80 gammatone filters than the gammatone filters of lower numbers, e.g. 20, which however is with similar number of filters and thus similar calculation speed. In other words, the Bark scale and the third-octave band filters have the similar performance on pitch analysis to the gammatone filters of high filter numbers for the sounds used in this study, but reduced computation time, since they consist of fewer bands. Between the Bark scale and the third-octave band filters, results based on which are similar, the calculation speed by the third-octave band filters is slightly quicker. Based on these results, therefore, the third-octave band filters is chosen for the pitch calculation of environmental sounds in this study. They can be seen as an approximation of auditory filters and a simplification of the high number filterbanks but with similar performance and accelerated computation speed.

In sum, among the pitch models in theory and application through literature, the temporal method may be applicable to pitch analysis of environmental sounds. The simplified temporal model is implemented with the third-octave band filters for the further pitch analysis in this study, based on the comparison among the values of the parameters and a number of filterbanks available for the optimizational performance.

5.6 Pitch Parameters Based upon Statistic Analysis

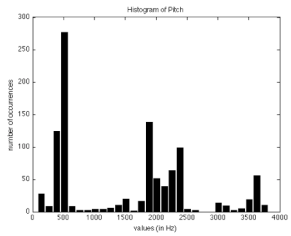
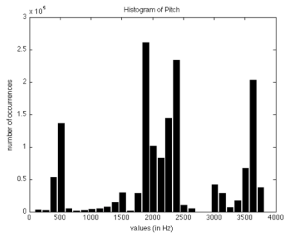
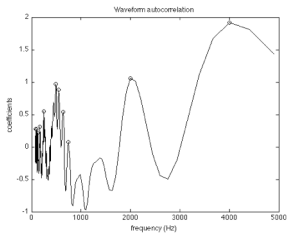
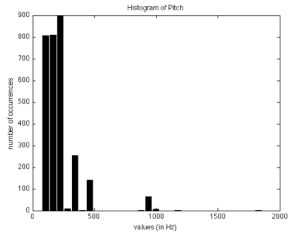
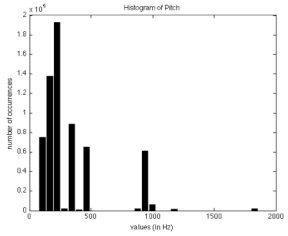
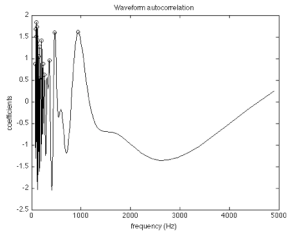
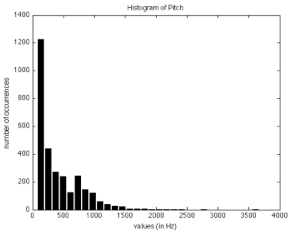
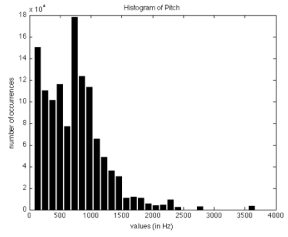
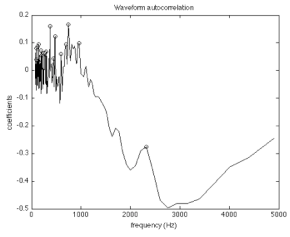
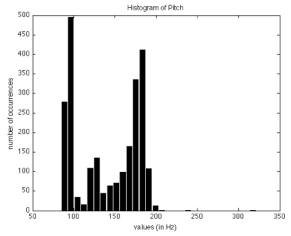
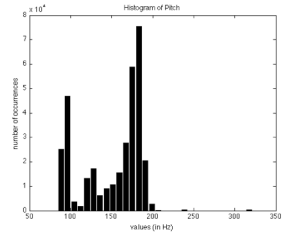
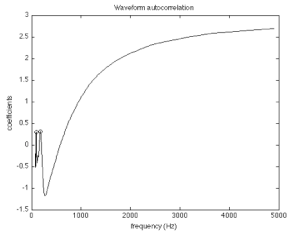
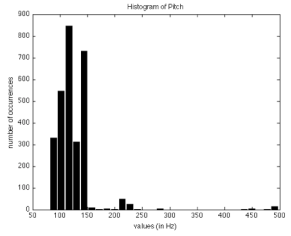
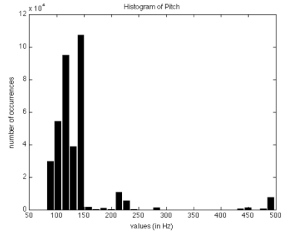
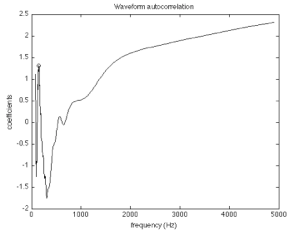
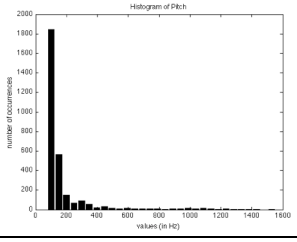
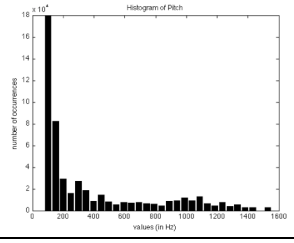
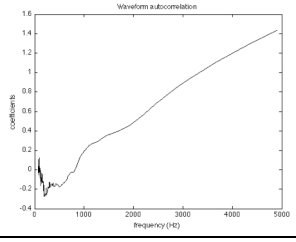
In order to describe the pitch characteristic of each sound, a number of basic statistic indices are calculated from the pitch results based on the algorithm selected in Section 5.5.

For variation of pitches over time, to simplify the calculation, only the most prominent pitch (the first pitch) is calculated in each successive frame. The histograms of the first pitches, shown in Table 5.6.1, generally do not differ much from those of the best four pitches. Thus, it is reasonable to think that the first pitches alone reflect the pitch characteristics over time to some degree. When also take into account the strength of each pitch, weighted histograms can be computed by adding the strength (amplitude) of the pitches instead of counting the number of pitches in each frequency range (bin). Both the histograms and weighted histograms are shown in Table 5.6.1. It can be seen that the pitch values are non-normally distributed along the linear frequency scale. To summarise the pitch data over time, therefore, a number of indices of basic descriptive statistics can be calculated from the data of each sound, which include median, mode (the value which occurs most frequently in the data), maximum, minimum, range (the difference between the maximum and minimum), and percentiles, in addition to mean and standard deviation. For the values of pitches over time of the 13 sounds, Table 5.6.2 shows mean, median (50% percentile), mode, standard deviation, maximum, minimum, range, and 5%, 10%, 25% (first quartile), 75% (third quartile), 90% and 95% percentiles. In addition, it shows the ratio of pitch over the duration, i.e. the ratio between the numbers of frames with pitch produced and the total frames in the duration. In terms of variation of pitch strength over time, similar indices are calculated, shown in Table 5.6.2.

Another way for describing the average pitches over the whole duration is to calculate the pitches from ACFs based on the whole duration, in contrast to those based on successive frames. Table 5.6.1 shows the SACFs for pitch computation. The values of best four pitches of the 13 sound recordings, as well as corresponding pitch strengths can be seen in Table 5.5.2. In Table 5.6.1, it can be seen that though the weighted histograms, in which each pitch is weighted by its pitch strength, and the corresponding characteristic have not been described by any index, the shapes of the SACFs are somehow similar to those of the weighted histograms. Thus, indices based on the SACF would to some degree reflect the characteristics of weighted histogram. For simplification, only indices based on SACF are extracted here, including the best four pitches, as well as corresponding pitch strengths.

Among all the indices of pitch value and strength shown in Table 5.6.2, some of them do not show any difference of the 13 sounds, e.g. minimum value of pitch, or generate results rather similar to each other, e.g. 5% and 10% percentiles. As a result, to reduce the number of indices, a number of indices are removed from the index set, indicated in the grey colour. The relevance and the correlations of the remaining indices are further analysed in Chapter 6.

Table 5.6.1 Pitch statistics of 13 sounds by different methods: histograms and SACF

	Histogram *p=mirpitch('folder','Klapuri','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0,'frame','mono'); mirhisto(p,'Number',30)*	Weighted Histogram *p=mirpitch('folder','Klapuri','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0,'frame','mono'); mirhisto(p,'Ampli','Number',30)*	SACF *[p,a]=mirpitch('folder','Klapuri','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0)*
Birdsong			
Church bells			
Fountain			
Machine			
Music			
River			

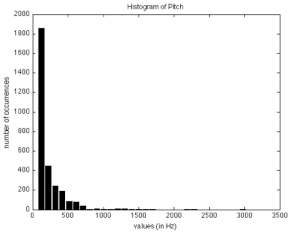
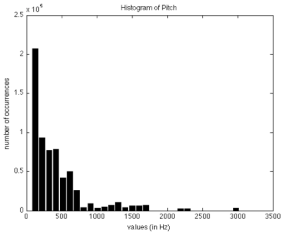
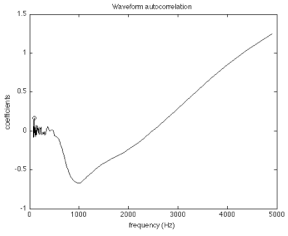
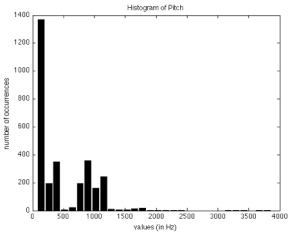
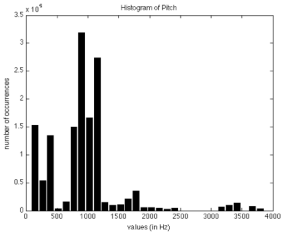
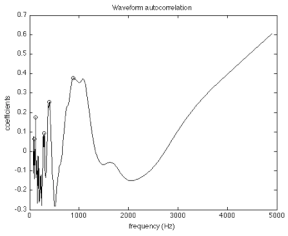
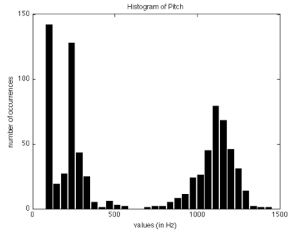
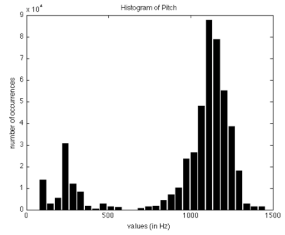
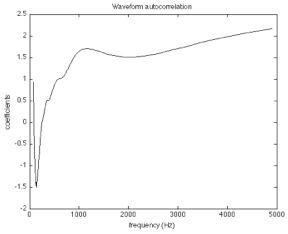
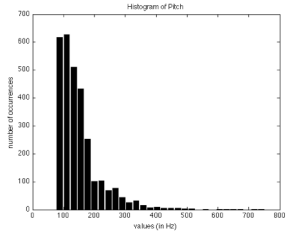
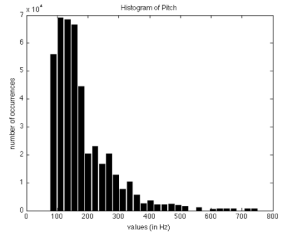
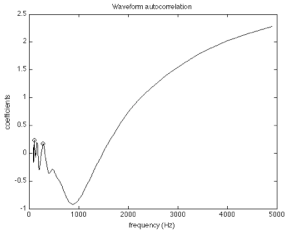
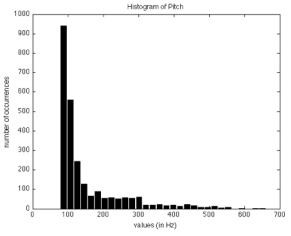
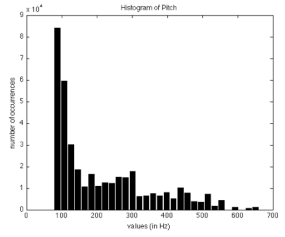
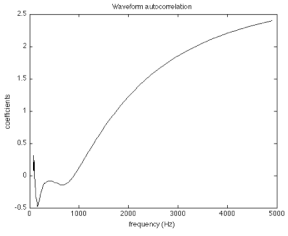
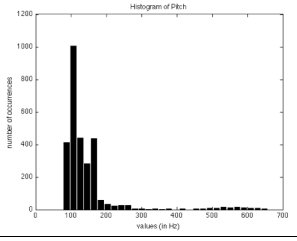
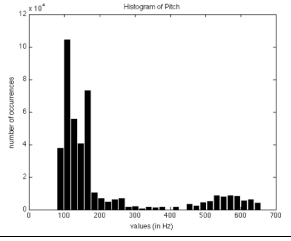
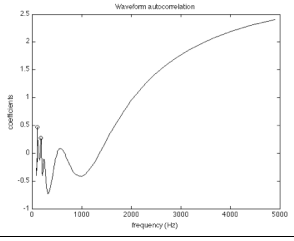
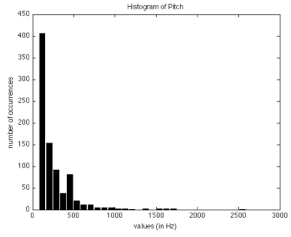
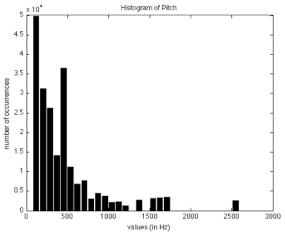
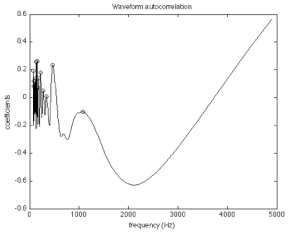
	Histogram	Weighted Histogram	SACF
Sea waves			
Stream			
Traffic			
Voice			
Wind			
Big Ben			
Victoria Park			

Table 5.6.2 Statistics of pitches and pitch strengths by frame of 13 sounds

		Birds ong	Chur ch bells	Foun tain	Mac hine	Musi c	River	Sea wave s	Strea m	Traffi c	Voic e	Wind	Big Ben	Victo ria Park
p=mirpitch('folder','Klapuri','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0,'frame','mono');	Count/2996	0.341	1.000	1.000	0.794	0.961	0.999	0.997	0.997	0.256	0.986	0.850	0.965	0.280
Pitch	Average	1501	212	406	141	123	173	216	477	636	150	151	145	257
pv=mirgetdata(p)	Median	1825	174	265	149	112	110	137	292	350	134	107	118	163
	MODE	470	170	90	90	110	90	90	120	240	90	90	100	150
	STDEV	1052	152	364	39	40	190	210	460	467	70	97	93	232
	STDEVA	940	152	364	67	46	190	210	460	364	71	104	96	168
	MIN	81	76	75	85	81	76	76	76	76	77	78	78	76
	MAX	3818	1857	3671	321	498	1553	3025	3860	1456	751	661	658	2588
	MAX-MIN	3738	1781	3595	237	417	1477	2949	3784	1379	675	583	580	2512
	Percentile.Exc 5	403	84	83	91	89	81	83	83	87	86	83	88	87
	Percentile.Exc 10	438	85	93	92	93	84	87	89	100	92	86	95	95
	Percentile.Exc 25	475	114	130	95	99	93	97	113	215	104	93	100	132
	Percentile.Exc 75	2254	222	602	178	146	151	247	860	1113	169	167	160	302
	Percentile.Exc 90	3031	352	918	185	147	302	442	1095	1194	236	290	176	464
	Percentile.Exc 95	3590	467	1099	187	149	520	592	1141	1238	285	376	277	638
Pitch strength	Average	0.952	0.565	0.180	0.290	0.939	0.220	0.263	0.248	0.368	0.450	0.254	0.373	0.501
pa=get(p,'Amplitude')	Median	0.827	0.554	0.172	0.268	0.887	0.216	0.255	0.240	0.359	0.413	0.250	0.358	0.388
	STDEV	0.376	0.196	0.044	0.147	0.351	0.066	0.071	0.068	0.227	0.186	0.114	0.131	0.311
	STDEVA	0.502	0.196	0.044	0.176	0.389	0.067	0.073	0.069	0.198	0.192	0.139	0.146	0.279
	MIN	0.538	0.117	0.090	-0.111	0.346	-0.123	-0.215	-0.198	-0.299	0.073	-0.220	-0.047	0.221
	MAX	5.958	1.219	0.457	1.345	4.609	0.620	0.713	0.601	0.934	1.728	0.786	1.017	2.979
	MAX-MIN	5.420	1.102	0.367	1.456	4.263	0.742	0.928	0.799	1.233	1.654	1.006	1.064	2.759
	Percentile.Exc 5	0.636	0.279	0.128	0.086	0.556	0.126	0.168	0.161	-0.031	0.231	0.074	0.187	0.268
	Percentile.Exc 10	0.665	0.317	0.136	0.130	0.609	0.146	0.186	0.176	0.101	0.260	0.127	0.219	0.287
	Percentile.Exc 25	0.712	0.407	0.150	0.196	0.708	0.177	0.216	0.203	0.223	0.323	0.184	0.276	0.325
	Percentile.Exc 75	1.067	0.704	0.200	0.365	1.052	0.261	0.300	0.284	0.535	0.534	0.321	0.451	0.583
	Percentile.Exc 90	1.411	0.812	0.236	0.465	1.381	0.302	0.354	0.332	0.666	0.688	0.393	0.548	0.816
	Percentile.Exc 95	1.638	0.890	0.262	0.554	1.564	0.334	0.392	0.371	0.737	0.789	0.442	0.617	1.059

It is noted that further analysis could be made based on variation of pitches over time and the SACFs from which pitches are calculated, and more indices could be extracted, e.g., variance of successive pitches, and pitch ambiguity – in addition to pitch strength indicated by absolute height of peak in the SACF pattern – which is thought to be related to the relative heights and number of neighbouring peaks in the pattern (Wightman 1973). However, it is not in the scope of the current study.

In sum, based on both pitch over time and average pitches for the whole duration, a number of descriptive statistic indices have been extracted from pitch feature of sound, including pitch value (height), strength, and ratio. For pitch value, the indices are mean, median, mode, standard deviation, range, and 5%, 25%, 75% and 95% percentiles of values of pitch over time, and values of the best four average pitches for the whole duration. For pitch strength, they are mean, median, standard deviation, range, and 5% and 95% percentiles of pitch strength over time, and strengths of best four average pitches for the whole duration. The index related to pitch ratio is the percentage of audible

pitches over time. These pitch indices are used for the analysis of characteristics of different types of sound in the next chapters.

The commands in Matlab with MIRtoolbox responding to the algorithms for the pitch computation are

p=mirpitch('folder','Klapuri','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0,'frame','mono');
for pitch over time, and

p=mirpitch('folder','Klapuri','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0,'Total',4);
for average pitches in the whole duration;

pv=mirgetdata(p)
pa=get(p,'Amplitude')
for accessing the data of value and strength of pitch respectively. The complete program is available in the accompanying CD-ROM.

5.7 Conclusions

Among the different pitch algorithms in literature, including temporal models and spectral models for pitch perception and simplified model for applications in music and speech, the temporal method is found to be applicable to the pitch analysis of environmental sounds, by examining their performances with 13 types of common sound in soundscapes.

Based on the temporal method, a simplified model is implemented with Matlab program. Through comparisons among the values of the parameters and a number of filterbanks available, the model is based on the calculation of decomposition through third-octave band filters, autocorrelation computation and pitch selection, with a number of parameters controlling the procedure. The pitch range calculated is between 75 and 5000 Hz.

A number of indices that describe the pitch feature of sound are extracted for the further pitch analysis of environmental sounds in this study. These indices include mean, median, mode, standard deviation, range, and percentiles of values and strengths of pitch over time, and values of the best four average pitches for the whole duration and their corresponding pitch strengths, as well as the percentage of audible pitches over time.

Chapter 6
Applicability of rhythm algorithms to environmental sounds

Rhythm is a general term that refers to the time-dependent properties of events or sound. This chapter explores the applicability of rhythm features and algorithms to environmental sounds. Here, among a number of specific rhythm concepts in music, e.g. beat, tempo, and meter, event (or note) detection models and tempo models are considered for environmental sounds. An event of music refers to a musical note (Brown 1993), while for environmental sounds, it refers to a salient pulse in signal in this study. Tempo models analyses the periodicity and its rate of sound. It is expected that some sound events in environmental sounds may not exhibit regular and periodic characteristics as most music pieces do, but certain types of environmental sounds may.

This chapter firstly explores the events models in Section 6.1, implemented and compared for environmental sounds. Then, in Section 6.2, a number of parameters/algorithms to analyse the events are derived from the systematic exploration of alternative methods proposed in literature, including event density and attack. Tempo or periodicity, also as a descriptor of events, can be calculated directly from temporal variation of signal without events data, discussed in Sections 6.3 and 6.4.

6.1 Events

For environmental sounds with single sound source in this study, events are assumed to be salient pulses in signal, which may respond to notes in music. An onset of event is defined as the instant of an attack of pulse, characterized by fast changes in intensity, pitch or timbre of sound (Bello *et al.* 2004), e.g. the beginning of note.

6.1.1 Event detection based on temporal pattern of total amplitude

Zwicker and Fastl (1999, p274-275) have presented a rhythm model that the rhythm of sound is calculated on the basis of temporal pattern of loudness. Basically, each maximum of the loudness-time function indicates a rhythmic event.

According to the model of Zwicker and Fastl (1999, p274-275), similar modelling is implemented in Matlab with MIRtoolbox in this section. The modelling consists of firstly

computation of the temporal pattern of amplitude of signal and next selection of the maxima of the pattern that indicate the events. Alternative to loudness, the temporal pattern of amplitude of signal can be computed by intensity, root-mean-square (RMS) (pressure) – the root of the mean of the square of the amplitude, or amplitude envelope of signal – the global outer shape outlining the extremes of amplitude (Lartillot *et al.* 2008).

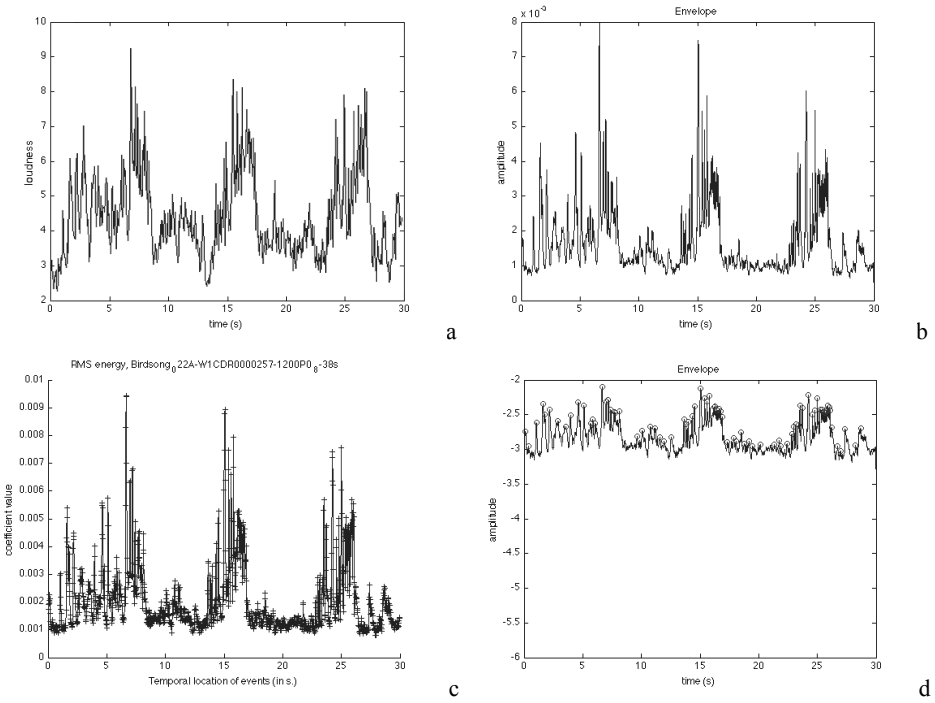


Figure 6.1.1 Loudness (a), envelope (b) and RMS of successive frames (c) of a birdsong recording, and event detection on envelope in logarithm scale (d), where (b), (c) and (d) corresponding to the commands as below respectively in Matlab:

mirenvelope('Birdsong')
mirrms('Birdsong','Frame')
e=mirenvelope('Birdsong','Log');
p=mirpeaks(e,'Threshold',0,'Contrast',0.03,'reso',0.05,'Chrono')

Figure 6.1.1 (a-c) show respectively the temporal patterns computed by the loudness, envelope, and RMS values of successive frames of signal, using a birdsong recording for illustration. An envelope can be estimated by a full-wave or half-wave rectification and a further smoothing of the signal. Here, the full-wave rectification is used, which converts all the negative lobes of signal into positive, leading to a series of positive half-wave lobes. The smoothing step has a low-pass characteristic so that it retains from the signal only long-term variation, removing all rapid oscillations (Viemeister 1979). It can be performed through either a low-pass filter or a temporal integrator which sums the energy occurring within a certain time interval or 'window', which two are equivalent in effect

(also see Section 6.1.2). Here, the low-pass infinite impulse response (IIR) filter (autoregressive) is used. Figure 6.1.1 (c) shows the temporal pattern of the birdsong recording computed through RMS values of successive frames of signal. Each successive frame used here has a frame length of 50ms and half overlapping. Since the computation of RMS involves time integration, no further smoothing is needed.

Although loudness algorithm is available and the loudness data of the sound recordings used in this study has been analysed in Chapter 4, the software package for loudness calculation (ArtemiS) is not directly compatible with Matlab environment. From Figure 6.1.1 (a-c), it can be seen that the curve calculated by loudness somehow differs with the other two but not much, while these other two curves are almost the same though via different approaches. In order to reduce computation complexity here, the envelope method is used following for the calculation of the temporal pattern of amplitude of signal. Between the two alternative methods which are similar, the selection is not crucial.

There are a number of potential ways of measuring the amplitude of envelope. The envelope can be expressed in linear, logarithm, or nonlinear μ-law compression scale (The constant μ compromises between a close-to-linear ($\mu < 0.1$) and a close-to-logarithmic ($\mu > 10^4$) transformation (Klapuri *et al.* 2006)). According to Weber's law, which states that just-noticeable difference in a stimulus is proportional to the magnitude of that stimulus, Weber fraction is roughly constant. In other words, for wideband noise, the smallest detectable change in intensity is approximately a constant fraction of the intensity of the signal. If the smallest detectable change is expressed in decibels, i.e. as the change in level, it is constant too, of a value of about 0.5-1dB. This holds from about 20 dB to about 100 dB above the absolute threshold (Moore 1997, p64). Thus, the envelope is presented in logarithm scale for detection of event components (events) here.

While the maxima of temporal pattern of amplitude of signal indicate rhythmic events, more specifically, a set of parameters is specified to control the peak selection. The model proposed by Zwicker and Fastl (1999, p274-275) postulates three conditions for a maximum to be considered as a rhythmic event: only if first, it lies above a relative loudness value, i.e., 0.43 (as used in Zwicker and Fastl's model) of the loudness of the highest maximum within a relevant time; second, it produces a significant increase in relative loudness, of sufficient height greater than 0.12 of the highest maximum; and third, it separates temporally greater than 120 ms from other rhythmic events, which means the maxima that are taken into account are more than 120 ms apart (Zwicker and Fastl 1999, p274-275). These three conditions are represented by three parameters in the implementing of model here, for peak picking operation in the amplitude function, which are "*Threshold*", "*Contrast*" and "*Reso*" respectively corresponding to the conditions above (also see Chapter 5). Here, they are set to the values of 0, 0.03 and 0.05

respectively, which means maxima that produce increments larger than 0.03 of the highest maximum in log scaled envelope and are more than 50ms apart are selected as rhythmic events. It is without the restriction of threshold of absolute or relative envelope amplitude, because the single source sounds used in this study are all audible over the duration and the changes over time are relatively small compared to the absolute amplitude of envelope in log scale, as can be seen in Table 6.1.1.

Figure 6.1.1 (d) shows the detection of events, illustrated with the same birdsong recording as in Figure 6.1.1 (a-c), using the modelling procedure of computation of amplitude envelope of signal in logarithm scale and selection of maxima of the envelope controlled by the parameters. The detected events are indicated by small circles. Table 6.1.1 shows the events detected of the 13 sound recordings described and used in Chapter 5 using this implemented model.

Whereas most rhythmic events (or event components) can be detected by means of variations in intensity or loudness versus time functions, variations in pitch and timbre can also sometimes lead to rhythmic events despite constant loudness (Zwicker and Fastl 1999, p275). In such cases, the variations in pitch or timbre can be calculated according to the respective models in the previous chapters. However, in that this results in different models for different types of sound, thus, in order to search for a unitive (general) model for all types of environmental sounds, it may be better using specific loudness versus time functions instead of the total loudness as the basis for rhythm calculating (Zwicker and Fastl 1999, p274-275). It accords with the perception of noticeable change, which has been described in Chapter 2 Section 2.3.10. Nevertheless, this loudness-based model forms the most simple and basic model of rhythm, and represents the basic concept of modelling of rhythmic events.

6.1.2 Event detection based on temporal pattern of specific amplitude in critical bands

Event detection based on temporal pattern of specific loudnesses in critical bands may find its basis in the modelling of temporal resolution of the auditory system (see Chapter 2 Section 2.3.10). Summarising several models of temporal resolution, they have the general form, which consists of a number of stages (Moore 1997, p160). The initial stage is a bandpass filtering, which reflects the action of the auditory filters in the inner ear. Secondly, each filter is followed by a nonlinear device, which may be thought of as crudely representing some aspects of the process of transduction from excitation on the BM to activity in the auditory nerve. The nonlinear device is either a rectifier, square-law

device or compressive nonlinearity (Viemeister 1979; Moore *et al.* 1988; Moore 1997). The output of nonlinear device is always positive that half-wave rectifier resembles the way that neural spikes tend to occur for a particular polarity of the stimulating waveform, and square-law device derives the instantaneous power at the output of the bandpass filter. Thirdly, a smoothing device is assumed to smooth the internal representation of auditory stimuli, i.e., it has the effect, on the output of the nonlinear device, of smoothing rapid fluctuations while preserving slower ones. The smoothing device is often referred to as a lowpass filter or a temporal integrator (see Section 6.1.1). Finally, the output of the smoothing device is fed to a decision device, which detects and compares the timing of events in different frequency channels.

The model discussed in the last section (Section 6.1.1) can be seen as a simplification of these temporal resolution models, in which only one filter band is used, or without bandpass filtering. The calculation of envelope or RMS corresponds to the second and third stages in the models above, and the process of selection of maxima of the temporal pattern is somehow similar to the final stage.

The implementation here based on the models of temporal resolution involves firstly a bank of nearly critical-band filters. The filterbank divides the signal into 21 non-overlapping bands, which are the same as those for pitch simulation in Chapter 5 and the loudness calculation (Chapters 2 and 3). The lowest three bandpass filters are of one-octave, and the remaining eighteen are third-octave, which together cover the frequencies from 44 Hz to 18 kHz. In the second stage, the output of each filter is full-wave rectified to simulate the nonlinearity. For the third stage of smoothing, a temporal integrator averages the rectified band signals within a 100-ms half-Hanning (raised cosine) window (Todd 1994; Scheirer 1998; Klapuri 1999), which performs much the same energy integration as the human auditory system, masking rapid fluctuations while preserving slower changes. In effect, the output of the temporal integrator (through the second and third stages) resembles the amplitude envelope of the output of the bandpass filter (the first stage). An example is giving in Figure 6.1.2, using a birdsong recording, the same as the one used in Figure 6.1.1 in the last section. Figure 6.1.2 (a) shows the output through the first three stages, i.e. the amplitude envelopes of the respective frequency channels.

In terms of the final stage, i.e. a decision device concerning how to detect the events in each frequency band and to combine the results across bands, as well as how to determine the amplitude of each event, the implementation here is according to Klapuri's (1999) music note onset detection system. In this system, event components are firstly detected in each band, with their time and intensity determined, and then the components are combined to yield events.

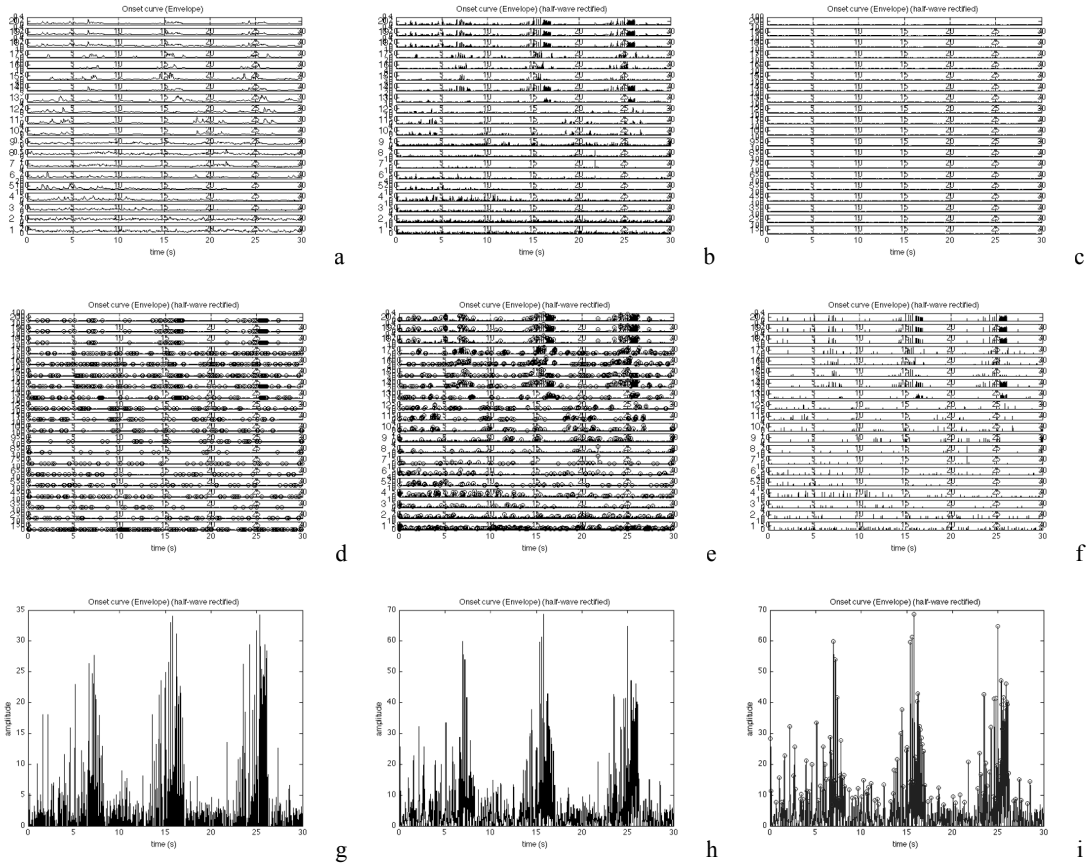


Figure 6.1.2 The procedure of implemented event detection model based on temporal pattern of specific amplitude in critical bands, illustrated with a birdsong recording. (a), (b), (c), (d), (e), (f), (g), (h), and (i) correspond to the commands as below respectively in Matlab:

(a) o=mironsets('Birdsong','Filter','FilterbankType','Klapuri','FilterType','HalfHann','Sum',0, 'PostDecim',180,'detect',0)
(b) o2=mirenvelope(o,'HalfwaveDiff')
(c) o=mirenvelope(o,'Log','HalfwaveDiff')
(d) p=mirpeaks(o,'Threshold',0.01,'Contrast',0,'NoBegin','Chrono')
(e) p2=mirpeaks(o2,'ScanForward',p,'Chrono')
(f) o=combinepeaks(p,p2,0.05)
(g) o=mirsum(o)
(h) o=mirenvelope(o,'Smooth',12)
(i) o=mirpeaks(o,'Total',Inf,'SelectFirst',0,'Threshold',0,'Contrast',0.1,'reso',0.05,'NoBegin','NoEnd', 'Order','Abscissa')

In each frequency band, instead of the points with maximum amplitude on envelope (as in the last section), the beginnings of discrete events, i.e. the onset of events, are to be detected. It is indicated by the maximum rising slope of amplitude envelope, and thus was practically calculated based on a half-wave rectified first-order difference function of the envelope, which is calculated by the differences between successive samples (Goto and Muraoka 1995; Scheirer 1998). In the function, the maxima (above a global threshold)

indicate the onset of events. Since psychoacoustically, perceived increase in signal amplitude is in relation to its level as discussed in the last section, event components are detected from a first-order relative difference function of envelope, i.e. first-order difference function of the logarithm amplitude envelope, with their onset time determined. In terms of the intensity of a detected event component, it is picked from the first-order difference function, determined by the maximum value between the onset to the point forward where amplitude stops increasing. After all event components in a band have been detected, with their onset time and intensities determined, only the components that separate temporally greater than 50 ms, i.e. the one with largest intensity within 50 ms are retained.

In the next step, event components from separate bands are combined to yield events of the overall signal. First, the event components from different bands are all sorted in time order, regarded as event candidates. Then, each event candidate is assigned an intensity value, which is calculated by collecting event components in a 50-ms time window around the candidate and adding their intensities together. From the event candidates, as a function of intensities vs. their times, the candidates are accepted as true events ones if their intensities (amplitude) are above a threshold and are more than 50 ms apart. Among candidates which are close to each others (within 50 ms), the loudest one is chosen.

The procedure of implemented model of event detection is illustrated in Figure 6.1.2, using a birdsong recording. Figure 6.1.2 (a) shows the envelopes of the 20 frequency channels computed through the first three stages. Figure 6.1.2 (b) and (c) respectively respond to the first-order difference functions and first-order relative difference functions of the envelopes, both of which are half-wave rectified. ("*HalfwaveDiff*" performs a half-wave rectification on the differentiation of the envelope.) (d), (e) and (f) respond to the detection of event components, determining their onset time from the first-order relative difference function and amplitudes from the first-order difference function. The parameters of "*Threshold*", "*Contrast*" and "*Reso*" for peak picking operation are set to the values of 0.01, 0 and 0.05 respectively, which means that the differences which are above a threshold of 0.01 of the maximum difference in log amplitude envelopes and are more than 50 ms apart are selected. In (g) and (h), the event components from different bands are combined. The "*Smooth*" operation smooths the event components function by averaging the components in a similar way to adding together amplitudes of event components in a time window. (i) responds to events selection from the combined event components function of overall signal. The parameters of "*Threshold*", "*Contrast*" and "*Reso*" for peak selection are set to the values of 0, 0.1 and 0.05 respectively.

This model follows the temporal resolution of human auditory system and is much

based on Klapuri's (1999) music note onset detection system. It detects events in different frequency bands separately and then combine results in the end, which differ from the model in the last section that process the amplitude envelope of signal as a whole.

6.1.3 Event detection based on spectral flux

Another approach for event detection processes the signal in the frequency domain using a Fourier transform, unlike the previous approaches as discussed in the last sections that process in the time domain based on energy of temporal waveform as a whole or in subband. Essentially, this approach is the same as the subband analysis approach in Section 6.1.2, but differ from it in processing procedure. While the approach in the last section uses firstly a filterbank decomposition of audio waveform and then compares the energy between time windows, this approach involves first a window (or frame) decomposition and then measures the distance or dissimilarity between spectra of frames, indicating events by large distances (Alonso *et al.* 2004; Bello *et al.* 2004). Nevertheless, both the approaches follow the principle of temporal resolution of the auditory system, however, the latter is superior computationally efficient to the former.

In this section, this approach of event detection is implemented in Matlab with MIRtoolbox using spectral flux method. Spectral flux is a measure of the changing of frequency content (spectrum) of a signal with respect to time (Laroche 2004), calculated by the distances between the power spectra of successive frames. The spectral flux of the k^{th} frame, SF(k), is expressed as

$$SF(k) = \sum_{i=0}^{n-1} s(k,i) - s(k-1,i)$$,

where s(k,i) is the value of the i^{th} frequency bin of the k^{th} frame; s(k-1,i) is that of the previous frame to the k^{th}. The spectral flux of the k^{th} frame is thus calculated by subtracting the values of each bin in the previous spectrum from those of corresponding bin in the current spectrum and summing up these differences. Spectral flux of a signal consists of those of successive frames as a function of frames or time, based on which events are detected.

The implementation here involves first a frame decomposition, of a frame length of 50 ms and hop factor of 0.5, i.e. with half-overlapping. In each frame, the spectrum of the short signal segment is computed by means of a Fourier transform, which transforms the temporal signal into frequency domain. Euclidian distance is used for measuring the difference between successive frames. In order to focus on increase of energy, only positive contributions of the frequency bins are summed (controlled by the parameter of

"*Inc*"). The peaks of the spectral flux of signal are selected as onset of events, based on the conditions as discussed in the last sections. Correspondingly, the parameters of "*Threshold*", "*Contrast*" and "*Reso*" are set to the values of 0, 0.2 and 0.05 respectively, i.e., the peaks that produce increments greater than the adjacent instants by 0.2 of the largest increment (and thus are above a threshold of at least 0.2 as well) and are more than 50ms apart are selected as events. The values of the parameters are determined by balancing the 13 recordings. An example can be seen in Figure 6.1.3, using a same birdsong recording as used in the last sections for illustration, where the detected events are indicated by small circles.

An alternative method consists in computing distances not only between strictly successive frames, but also between all frames or instants of a signal within a temporal neighbourhood of pre-specified width (Foote and Cooper 2003; Lartillot *et al.* 2008). If inter-frame distances between all possible pairs of frames (without the restrict of temporal width) are computed, the result can be embedded into a two-dimensional representation, which is termed a similarity matrix (Foote and Cooper 2003) (also described in Section 6.3.3). Onsets of events can be derived from the main diagonal of the similarity matrix by a matched-filter approach. That is correlating a Gaussian-tapered checkerboard kernel along the main diagonal of the similarity matrix in a pre-specified width. The correlations, referred to as novelty score (Foote and Cooper 2003), which are time-indexed computed by the convolution, form a novelty curve. Large peaks in the novelty curve indicate the positions of transitions along the temporal variation of spectral distribution, i.e. the onsets of events.

An example illustrated with the same birdsong recording is shown in Figure 6.1.3. Here, firstly, the signal is frame decomposed with a frame length of 50ms and half overlapping. Then, the spectrum of the signal in each frame is computed using a Fourier transform, the same as in the method of spectral flux. Thirdly, the distance or dissimilarity between the spectra is measured for all frame combinations within a temporal width of 64 samples, i.e. 1.6 s. The distance is calculated by cosine similarity (Foote and Cooper 2003), and thus is normalized to be independent of magnitude. Then, the value of distance or dissimilarity between each pair of frames is transformed into the value of similarity by an exponential function. A novelty curve is derived by a convolution along the main diagonal of the similarity matrix with a Gaussian checkerboard kernel. Finally, the peaks in the novelty curve that produce increments greater than the adjacent instants by 0.15 of the largest increment and are more than 50ms apart and are selected as onsets of events. The onsets of events of the 13 sound recordings detected using this procedure are shown in Table 6.1.1.

Additionally, in this method, other features can be used to characterize the content of

frames in addition to spectrum particularly, e.g. mel-frequency cepstral coefficient (MFCC), or psychoacoustic attributes such as loudness, pitch or timbre. In this way, it would be similar to that in Section 6.1.1, if using loudness or energy as the parameterization.

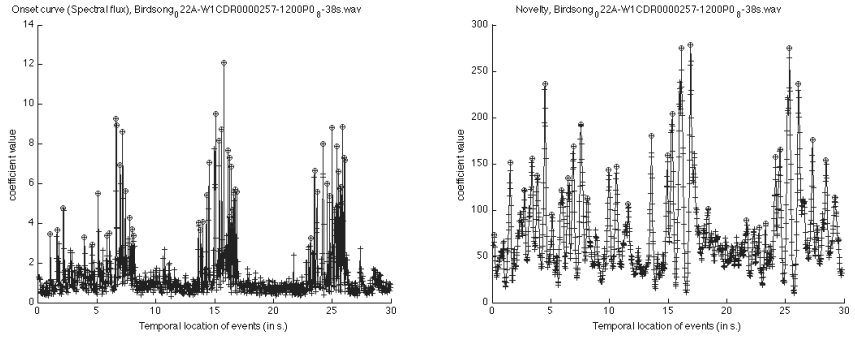


Figure 6.1.3 Event detection based on the computation of spectral flux (left) and novelty (right), illustrated with a birdsong recording, corresponding to the commands as below in Matlab:

mironsets('Birdsong','SpectralFlux','Threshold',0,'Contrast',0.2)
n=mirnovelty('Birdsong','Normal',0);
p=mirpeaks(n,'Contrast',0.15,'reso',0.05,'Chrono')

6.1.4 Comparison of event detection models

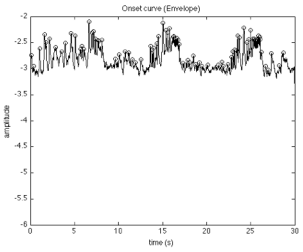
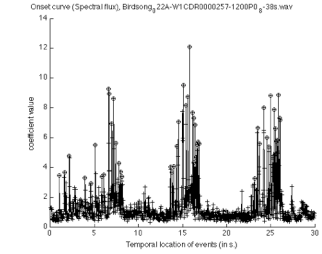
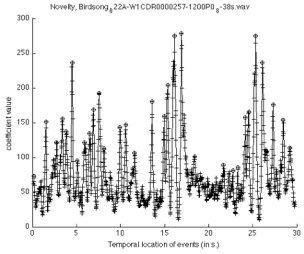
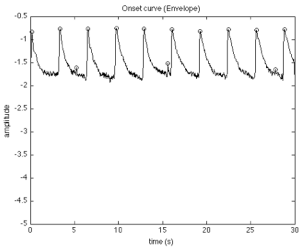
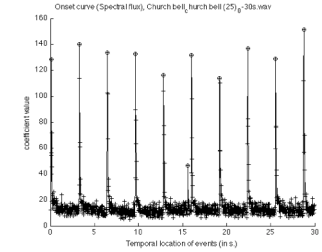
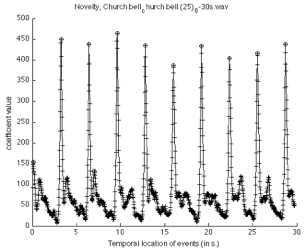
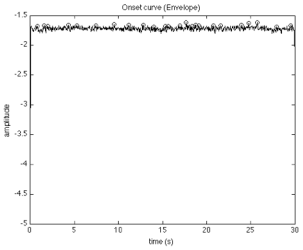
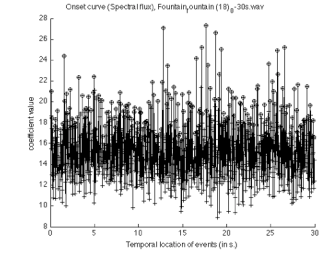
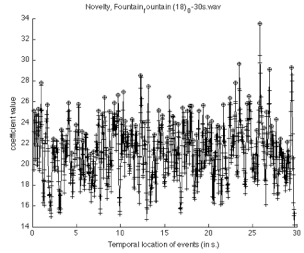
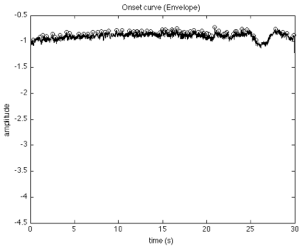
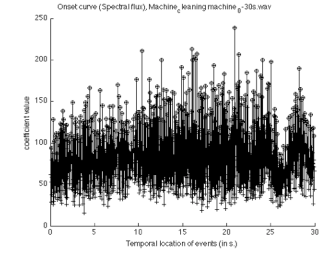
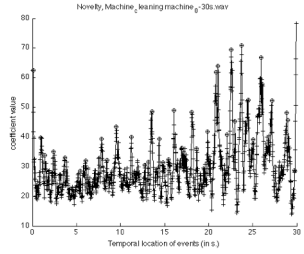
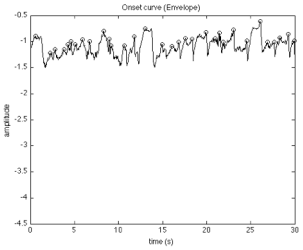
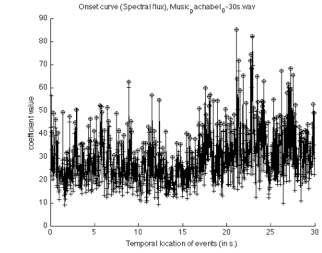
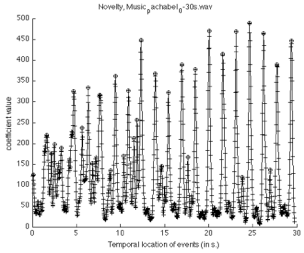
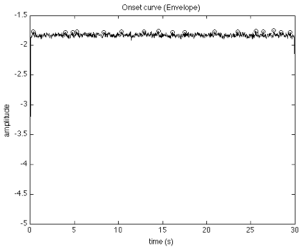
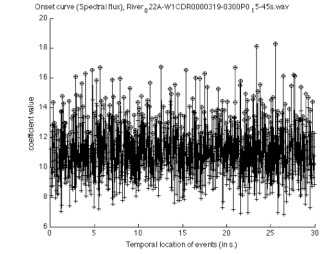
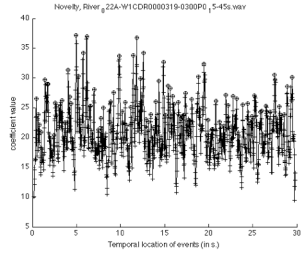
In brief, although detection of events in a signal can be computed by various methods as discussed above, generally, it is based on firstly computation of a temporal curve which contains the information for events, and then an operation of peak picking performed on it (Lartillot *et al.* 2008). This event detection curve can be time functions of energy (amplitude envelope), timbre or pitch of audio signal, or of energy of frequency bands – similar to spectral content as well (Foote and Cooper 2003), or be difference functions of these features (Klapuri 1999) which focus on the onset of events. The peaks of the event detection curve correspond to the loudest positions or onsets of events.

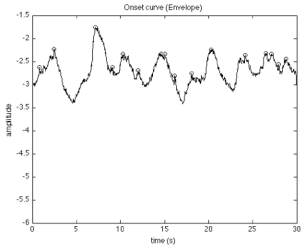
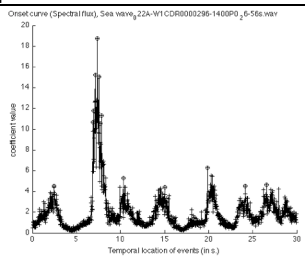
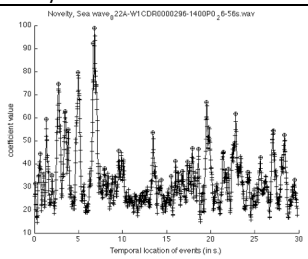
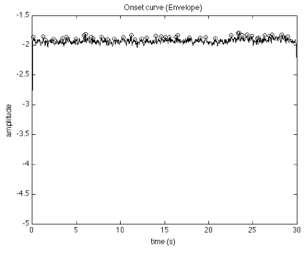
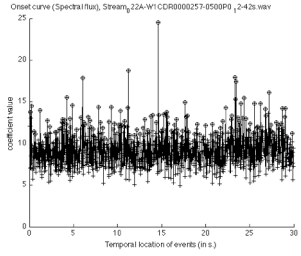
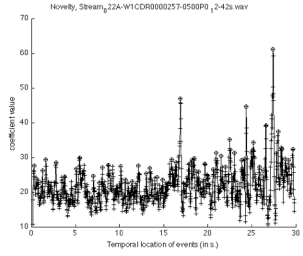
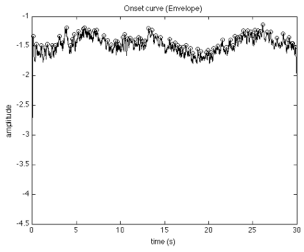
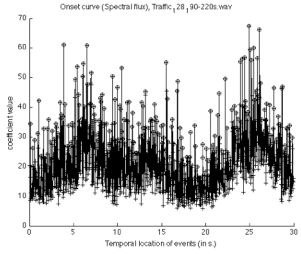
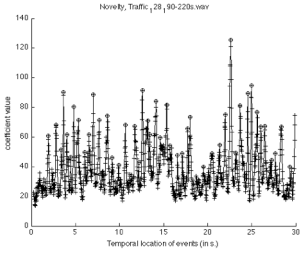
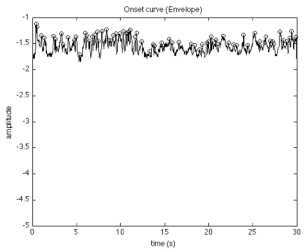
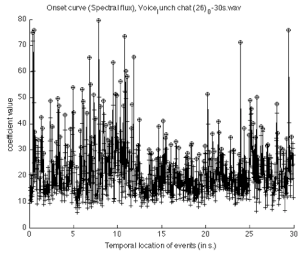
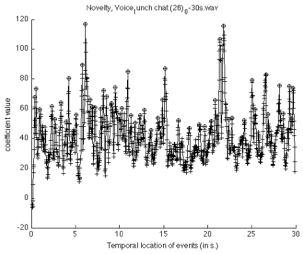
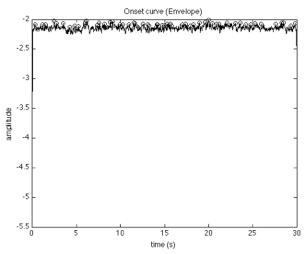
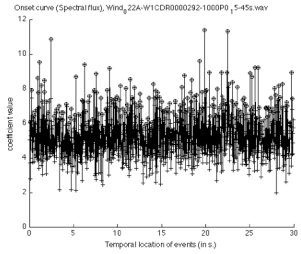
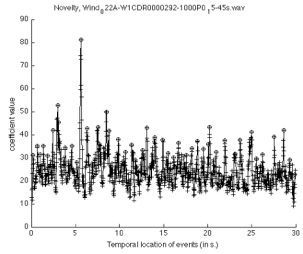
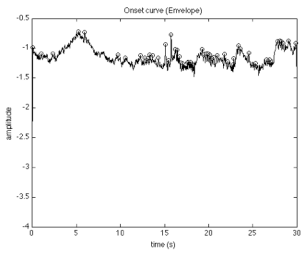
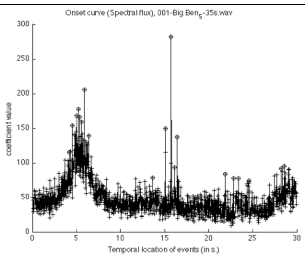
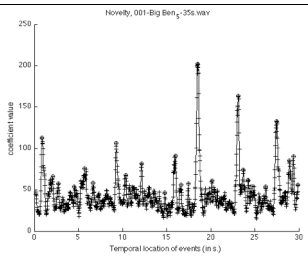
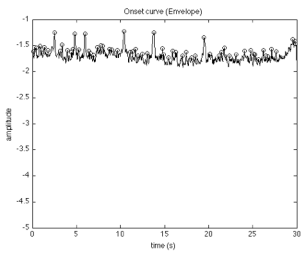
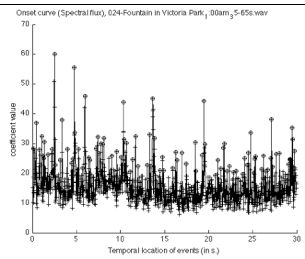
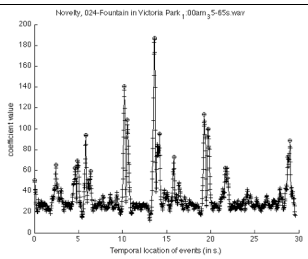
Table 6.1.1 shows the events of the 13 sample recordings detected by the different methods as described in the last sections (three among the four methods are shown as discussing following), indicated by small circles. For each method, the parameters are set to have the optimized results for all the 13 recordings. It can be seen by comparison among the results that the event results that may somehow differ based on different methods, since the focuses of these methods slightly differ.

The first method, the one based on total amplitude envelope, detects events by the overall energy or loudness, whereas the other three methods, based on a number of envelopes in frequency bands, spectral flux, or similarity, also take into account the influence of changes in spectral content on detection. Also, the first method detects the positions of maximum energy of events, while the others emphasise the onsets of events. Between the two methods that based on envelopes in frequency bands and on spectral flux, they have the similar principles but with different processing procedures. Since the former is much more computationally complex and thus more difficult to control, the latter would be used instead for event detection, and the results by the former are not shown in the table. Different from these two methods, the method based on novelty detects events not depending on the instantaneous changes in signal, but the changes that last for a longer time. For example, vibrato in music and repeating birdsongs would be detected as single events, which would however be considered to be multiple if using methods based on envelope or spectral flux due to high variability in energy and pitch (Lartillot *et al.* 2008).

From the results of the 13 recordings, shown in Table 6.1.1, it can be seen that for signals with definite pulses (or events), e.g. the church bells, all the methods detect them well. For the music recording, both the methods that based on overall envelope and on novelty detect the notes well, better than the method based on spectral flux. For sounds which are relatively stable in energy variation, such as the recordings of fountain, river, stream, and wind, those two methods (based on overall envelope and on novelty) detect smaller numbers of events than the method based on spectral flux does. It may result from that, differing from the method based on spectral flux, the one based on overall envelope ignores small changes of energy in frequency bands or spectral content, and the one based on novelty detects changes of relatively long durations. However, events in these sounds are difficult to identify according to the subjective hearing judgement of the author. For the birdsong recording, as discussed above, the methods based on envelope and spectral flux detect several repeating birdsongs as a number of individual events, while that based on novelty detect them as a single event. According to these results and discussions, both the methods based on overall envelope and spectral flux are used for event detection in this study, in order to focus on instantaneous changes or salient pulses in signal for single sounds. It may be expected that the method based on novelty would be useful for detection of sound source or action events in soundscape sounds. The differences of the results of event detection among the 13 sounds and among the three methods are further discussed in the next section.

Table 6.1.1 Event detection of 13 sounds based on different methods

	Envelope *o=mironsets('folder','Filter','Filterbank',0,'log','Contrast',0.03)*	Spectral Flux *o=mironsets('folder','SpectralFlux','Contrast',0.2)*	Novelty *n=mirnovelty('folder','Normal',0); p=mirpeaks(n,'Contrast',0.15,'reso',0.05,'Chrono')*
Birdsong			
Church bells			
Fountain			
Machine			
Music			
River			

	Envelope	Spectral Flux	Novelty
Sea waves			
Stream			
Traffic			
Voice			
Wind			
Big Ben			
Victoria Park			

6.2 Event Parameters

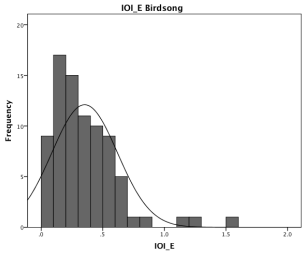
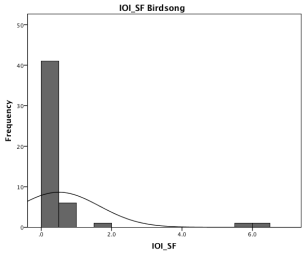
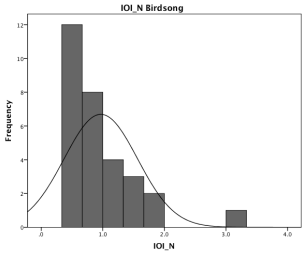
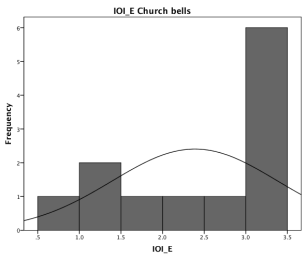
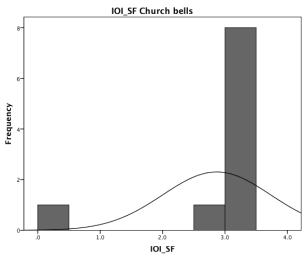
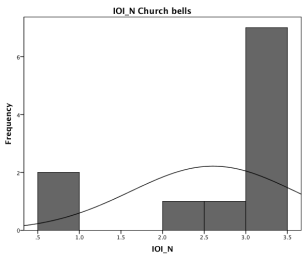
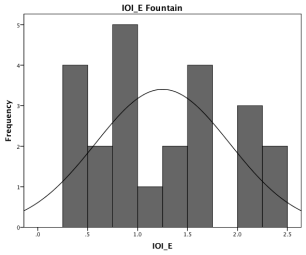
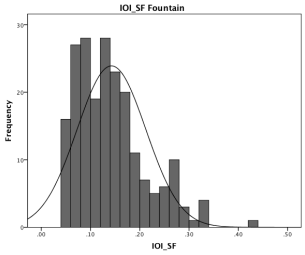
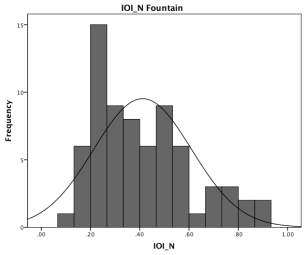
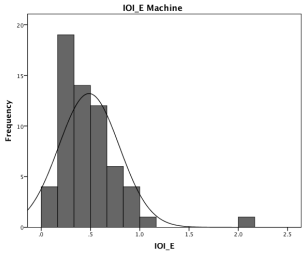
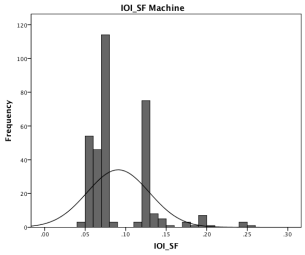
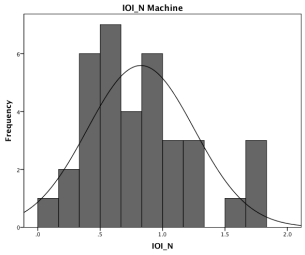
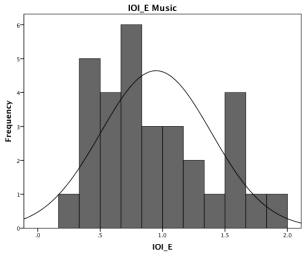
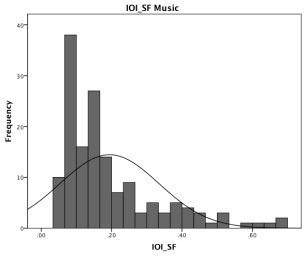
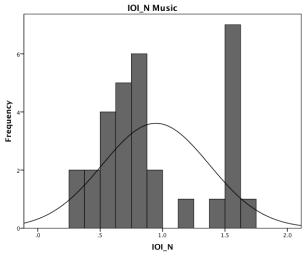
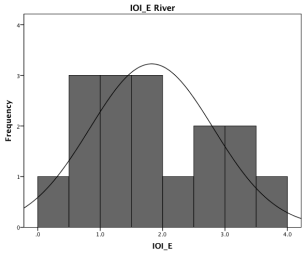
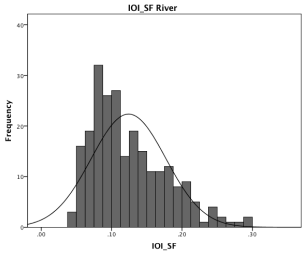
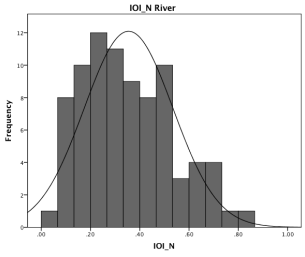
To describe the events detected by the methods above, a number of descriptors or parameters are developed, based on a systematic exploration of those proposed in literature (Lartillot *et al.* 2008). They are derived from the events detected and events detection curve that computed for the detection, which include inter-onset interval (IOI), event density, attack slope and periodicity. Each of the parameters is explained specifically in the following sections, and a series of basic statistic indices are calculated for each of the parameters with the software package of Excel or PASW Statistics.

While periodicity can be calculated either based on the detected events and event detection curve, or based on signal directly with specific methods, the computation of periodicity is discussed individually in Sections 6.3 and 6.4.

6.2.1 Event interval

Intervals between successive events, termed as inter-onset interval (IOI) in music analysis, are calculated from the results of event time or event onset time. The histograms of the event intervals calculated of the 13 samples based on the three methods are shown in Table 6.2.1. It can be seen that the distributions of the event intervals are non-normal for almost all the recordings, in other words, the normal curves do not fit histograms. For example, for church bells, the event intervals are mainly concentrated in some value ranges. Thus, a number of basic, descriptive, statistic indices are used to summarise the results of intervals, which include median, mode, maximum, minimum, range, and percentiles, in addition to mean and standard deviation. Table 6.2.2 shows mean, median, mode, standard deviation, maximum, minimum, range, and 5%, 10%, 25%, 75%, 90% and 95% percentiles of event intervals of each of the 13 sounds over the entire duration, for the different methods. Again, the results are different to some degree based on the different methods. Among these indices, 5% percentiles of both the methods do not show any value for some of the recordings, e.g. for church bells and sea waves, due to the small numbers of events detected in these recordings. As a result, these indices are removed from the index sets. In order to further reduce the number of indices that used for analysis in the next chapters, the correlations between the indices are represented graphically in Figure 6.2.1, from which the representative indices are remained. The remaining indices are indicated in the black colour in Table 6.2.2.

Table 6.2.1 Histograms of event interval (IOI histograms) of 13 sounds based on three different methods. The bin widths are not adjusted due to the automatic generation of Matlab.

	Envelope *o=mironsets('folder','Filter','Filterbank',0,'log','Contrast',0.03);* *ot=mirgetdata(o)* *ot{1,:}*	Spectral Flux *o=mironsets('folder','SpectralFlux','Contrast',0.2);* *ot=mirgetdata(o)* *ot{1,:}*	Novelty *n=mirnovelty('folder','Normal',0);* *p=mirpeaks(n,'Contrast',0.15,'reso',0.05,'Chrono')* *ot=mirgetdata(p)* *ot{1,:}*
Birdsong			
Church bells			
Fountain			
Machine			
Music			
River			

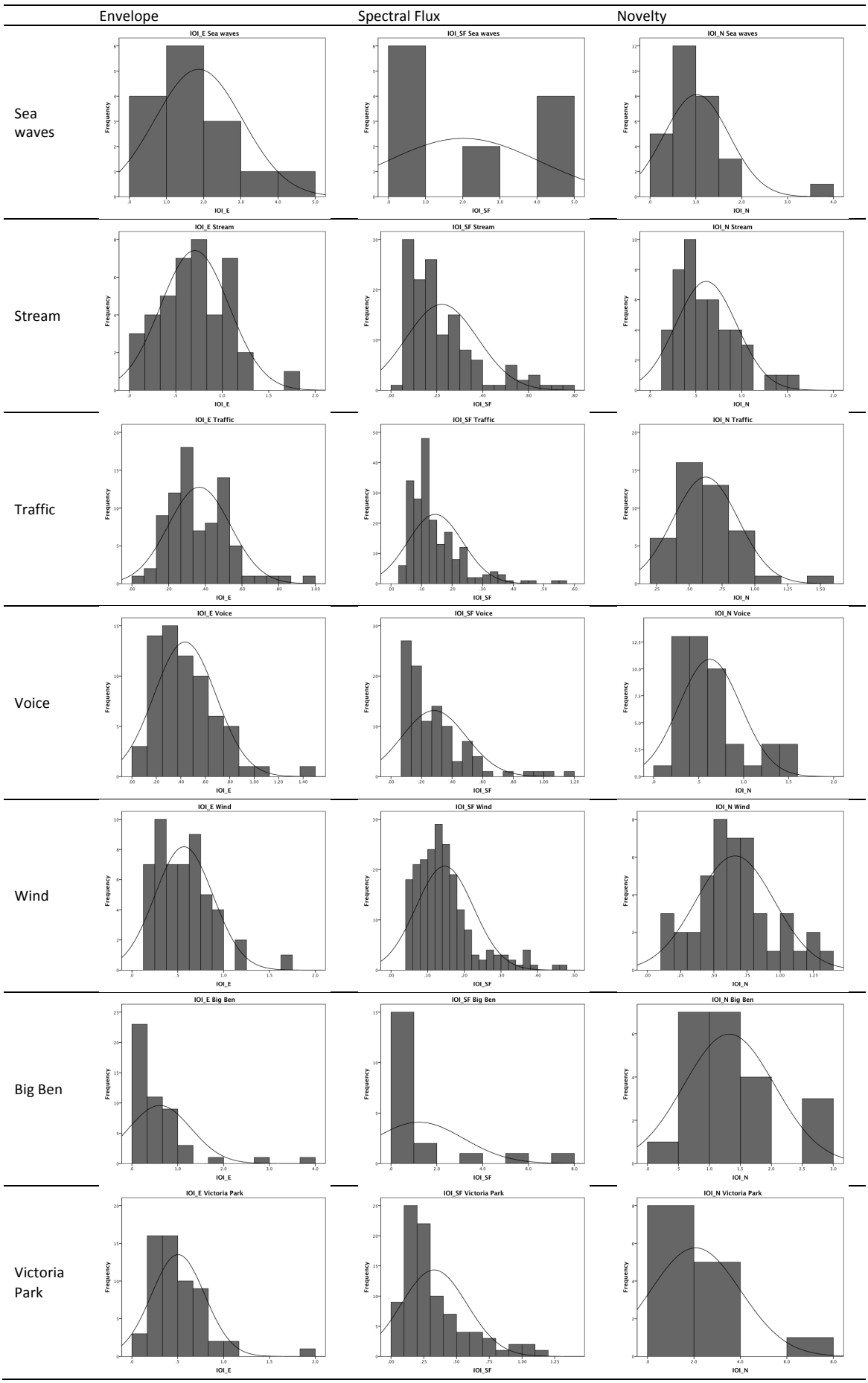

Envelope
Spectral Flux
Novelty
Sea waves
Stream
Traffic
Voice
Wind
Big Ben
Victoria Park
IOI_E Sea waves
IOI_SF Sea waves
IOI_N Sea waves
IOI_E Stream
IOI_SF Stream
IOI_N Stream
IOI_E Traffic
IOI_SF Traffic
IOI_N Traffic
IOI_E Voice
IOI_SF Voice
IOI_N Voice
IOI_E Wind
IOI_SF Wind
IOI_N Wind
IOI_E Big Ben
IOI_SF Big Ben
IOI_N Big Ben
IOI_E Victoria Park
IOI_SF Victoria Park
IOI_N Victoria Park
Frequency
IOI_E
IOI_SF
IOI_N

Table 6.2.2 Statistics of event interval of 13 sounds based on different methods

IOI		Birdsong	Church bells	Fountain	Machine	Music	River	Sea waves	Stream	Traffic	Voice	Wind	Big Ben	Victoria Park
Envelope	Average	0.352	2.388	1.250	0.485	0.947	1.825	1.865	0.705	0.368	0.434	0.568	0.609	0.503
o=mironsets('folder',	Median	0.288	2.919	1.200	0.454	0.826	1.617	1.712	0.679	0.322	0.387	0.542	0.379	0.442
'Filter','Filterbank',0,	Mode	0.2	3.2	1.6	0.3	0.7	0.8	1.7	1.1	0.3	0.3	0.3	0.2	0.3
'log','Contrast',0.03);	STDEV	0.267	0.995	0.674	0.307	0.444	0.989	1.180	0.368	0.169	0.253	0.317	0.679	0.291
ot=mirgetdata(o)	MIN	0.062	0.501	0.322	0.066	0.189	0.496	0.431	0.134	0.065	0.088	0.130	0.075	0.109
ot{1,:}	MAX	1.518	3.233	2.406	2.150	1.896	3.672	4.687	1.794	0.986	1.435	1.665	3.770	1.889
	MAX-MIN	1.456	2.733	2.084	2.084	1.707	3.176	4.256	1.660	0.921	1.347	1.535	3.695	1.780
	Percentile.Exc 5	0.063	-	0.326	0.130	0.276	-	-	0.158	0.135	0.133	0.146	0.102	0.151
	Percentile.Exc 10	0.072	0.651	0.350	0.196	0.440	0.700	0.560	0.175	0.168	0.170	0.177	0.179	0.219
	Percentile.Exc 25	0.173	1.441	0.686	0.325	0.635	0.904	0.852	0.434	0.256	0.253	0.312	0.212	0.301
	Percentile.Exc 75	0.465	3.186	1.678	0.590	1.240	2.636	2.337	0.990	0.491	0.569	0.736	0.812	0.648
	Percentile.Exc 90	0.659	3.222	2.281	0.833	1.594	3.448	4.177	1.112	0.556	0.773	0.960	1.268	0.758
	Percentile.Exc 95	0.871	-	2.391	0.965	1.762	-	-	1.300	0.688	0.886	1.242	2.313	1.005
Spectral Flux	Average	0.501	2.867	0.143	0.091	0.194	0.124	2.011	0.222	0.145	0.282	0.146	1.220	0.329
o=mironsets('folder',	Median	0.202	3.184	0.129	0.073	0.147	0.107	1.336	0.182	0.121	0.216	0.127	0.363	0.253
'SpectralFlux',	Mode	0.07	3.19	0.15	0.07	0.08	0.10	0.15	0.10	0.10	0.10	0.10	0.30	0.15
'Contrast',0.2);	STDEV	1.153	0.866	0.070	0.038	0.141	0.053	2.063	0.157	0.090	0.211	0.078	1.948	0.250
ot=mirgetdata(o)	MIN	0.052	0.432	0.045	0.048	0.053	0.045	0.122	0.048	0.046	0.075	0.047	0.155	0.053
ot{1,:}	MAX	6.292	3.193	0.431	0.251	0.699	0.296	4.821	0.750	0.572	1.144	0.472	7.249	1.178
	MAX-MIN	6.240	2.762	0.387	0.202	0.646	0.250	4.699	0.702	0.526	1.069	0.425	7.094	1.126
	Percentile.Exc 5	0.054	-	0.057	0.054	0.063	0.055	-	0.057	0.051	0.091	0.054	0.156	0.075
	Percentile.Exc 10	0.059	0.664	0.070	0.056	0.072	0.068	0.129	0.074	0.056	0.099	0.069	0.173	0.099
	Percentile.Exc 25	0.124	3.070	0.090	0.067	0.089	0.081	0.156	0.102	0.084	0.127	0.096	0.243	0.152
	Percentile.Exc 75	0.374	3.189	0.177	0.125	0.250	0.156	4.361	0.282	0.178	0.371	0.174	1.105	0.424
	Percentile.Exc 90	0.655	3.193	0.249	0.129	0.411	0.203	4.795	0.488	0.248	0.540	0.260	5.276	0.720
	Percentile.Exc 95	3.479	-	0.275	0.166	0.510	0.221	-	0.581	0.329	0.723	0.320	7.160	0.993
Novelty	Average	0.968	2.605	0.412	0.826	0.946	0.355	1.015	0.612	0.624	0.624	0.660	1.320	2.070
n=mirnovelty('folder',	Median	0.789	3.182	0.374	0.761	0.779	0.306	0.847	0.526	0.608	0.547	0.624	1.165	1.795
'Normal',0);	Mode	0.6	3.2	0.2	0.9	0.8	0.2	1.1	0.4	0.5	0.4	0.5	0.6	0.4
p=mirpeaks(n,	STDEV	0.596	0.990	0.198	0.428	0.428	0.180	0.712	0.332	0.249	0.344	0.296	0.733	1.939
'Contrast',0.15,'reso',	MIN	0.382	0.625	0.107	0.129	0.356	0.057	0.153	0.127	0.234	0.158	0.105	0.456	0.265
0.05,'Chrono');	MAX	3.273	3.189	0.908	1.782	1.652	0.836	3.883	1.575	1.511	1.508	1.369	2.974	7.353
ot=mirgetdata(p)	MAX-MIN	2.891	2.564	0.802	1.654	1.295	0.779	3.730	1.449	1.277	1.350	1.264	2.518	7.088
ot{1,:}	Percentile.Exc 5	0.384	-	0.153	0.195	0.367	0.109	0.208	0.152	0.262	0.239	0.145	0.465	-
	Percentile.Exc 10	0.458	0.646	0.186	0.345	0.463	0.131	0.334	0.263	0.362	0.282	0.274	0.529	0.315
	Percentile.Exc 25	0.606	2.437	0.248	0.484	0.610	0.216	0.554	0.370	0.426	0.370	0.461	0.637	0.511
	Percentile.Exc 75	1.209	3.188	0.525	1.072	1.513	0.476	1.322	0.821	0.752	0.764	0.810	1.828	3.158
	Percentile.Exc 90	1.778	3.189	0.747	1.571	1.572	0.623	1.703	1.048	0.961	1.258	1.113	2.623	5.594
	Percentile.Exc 95	2.562	-	0.826	1.765	1.623	0.703	2.895	1.346	1.080	1.423	1.282	2.921	-

Also, from Table 6.2.2 and Figure 6.2.1, it can be seen roughly that the recordings such as church bells, sea waves and Big Ben have higher range and 90% percentile values of event interval than the other recordings, i.e., some successive events in these recordings generally exhibit relatively large intervals. The specific and statistic characteristics of different types of sound in terms of event interval, as well as the relevance and correlations of these remaining indices, are further analysed intensively in Chapter 7.

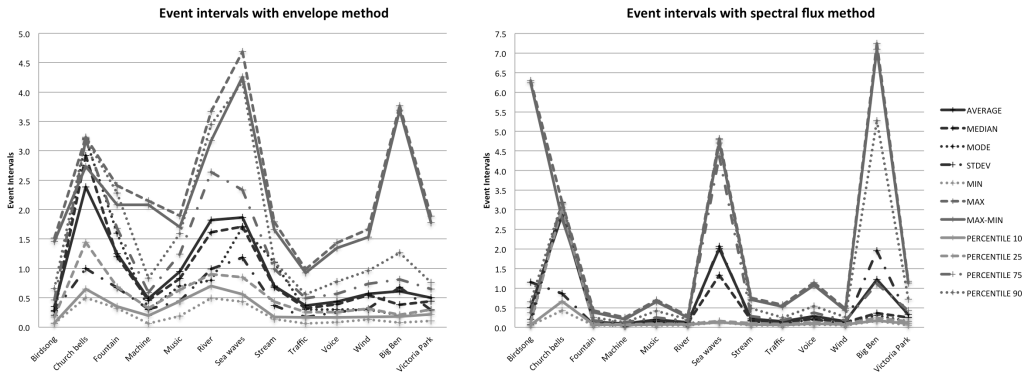


Figure 6.2.1 Statistics of event intervals of 13 sounds based on envelope method (left) and spectral flux method (right) of event detection

6.2.2 Event density

In addition to the parameter of event interval, event density, i.e. the average number of events per unit time (the unit of second is used here), estimates the frequency of occurrence of events. It is expected that event density is closely related to event interval, as they are in a reciprocal relationship. In Table 6.2.3, the event densities of the 13 sounds are shown, calculated through frequency of events over the whole duration – i.e. the number of events divided by the whole duration, and through the reciprocal of average value of event interval. It can be seen that the results by these two methods are almost the same, as expected. Thus, between these two, the frequency of events is used for calculation of event density in the analysis in the next chapters.

In addition to event density for the whole duration, variation of event density with time can be estimated. The variations of event density of the 13 sounds are computed based on the envelope method of event detection. An example can be seen in Figure 6.1.3 (left), using a birdsong recording. Here, for the successive windows or frames that used for calculation of variation, frame length of 3 seconds and hop factor of 0.1 (90% overlapping) are used. A number of basic statistic indices are used to describe the results of variation of event density, which include mean, median, mode, standard deviation, maximum, minimum, range, and percentiles, as shown in Table 6.2.3. It can be seen that the mean, median, and mode results are somehow similar, all of which are close to those calculated through event density of the whole duration. From the other indices, standard deviation, range, and 5% and 95% percentiles can be drawn to characterise the index set. The results show that birdsong and Big Ben recordings have higher standard deviation and range values than the others, which suggest that these sounds have relatively larger variations of event density over time.

Table 6.2.3 Average event density based on different event detection methods and statistics of variation of event density with time

ED		Birdsong	Church bells	Fountain	Machine	Music	River	Seawaves	Stream	Traffic	Voice	Wind	Big Ben	Victoria Park
Envelope *o=mironsets('folder','Filter','Filterbank',0,'log','Contrast',0.03);* *ot=mirgetdata(o)* *ot{1,:}*	(Count+1)/30	2.73	0.43	0.80	2.07	1.07	0.57	0.53	1.40	2.73	2.30	1.77	1.67	2.00
	1/Average	2.84	0.42	0.80	2.06	1.06	0.55	0.54	1.42	2.72	2.31	1.76	1.64	1.99
Spectral Flux *o=mironsets('folder','SpectralFlux','Contrast',0.2);* *ot=mirgetdata(o)* *ot{1,:}*	(Count+1)/30	1.70	0.37	7.00	10.97	5.13	7.97	0.43	4.50	6.90	3.50	6.80	0.70	3.03
	1/Average	2.00	0.35	7.01	11.02	5.16	8.06	0.50	4.51	6.92	3.55	6.83	0.82	3.04
Novelty *n=mirnovelty('folder','Normal',0);* *p=mirpeaks(n,'Contrast',0.15,'reso',0.05,'Chrono');* *ot=mirgetdata(p)* *ot{1,:}*	Count/30	1.00	0.37	2.37	1.20	1.03	2.73	0.97	1.60	1.47	1.57	1.50	0.73	0.47
	1/Average	1.03	0.38	2.43	1.21	1.06	2.82	0.99	1.63	1.60	1.60	1.52	0.76	0.48
Envelope *o=mironsets('folder',* *'Filter','Filterbank',0,* *'log','Contrast',0.03,* *'frame');* *ed=mireventdensity(o)* *mirgetdata(ed)*	Average	2.73	0.39	0.64	1.96	0.93	0.36	0.47	1.25	2.58	2.16	1.62	1.56	1.89
	Median	2.33	0.33	0.67	2.00	1.00	0.33	0.33	1.33	2.67	2.00	1.67	1.67	2.00
	Mode	1.67	0.33	0.33	1.67	1.00	0.33	0.33	1.00	2.33	2.00	1.67	2.00	1.67
	STDEV	1.22	0.17	0.33	0.64	0.33	0.25	0.28	0.38	0.44	0.66	0.37	1.04	0.43
	MIN	1.00	0.00	0.33	0.67	0.33	0.00	0.00	0.33	1.67	1.00	0.67	0.00	1.00
	MAX	6.00	0.67	1.67	3.00	1.67	1.00	1.33	2.33	3.67	3.33	2.67	4.00	2.67
	MAX-MIN	5.00	0.67	1.33	2.33	1.33	1.00	1.33	2.00	2.00	2.33	2.00	4.00	1.67
	Percentile.Exc 5	1.33	0.00	0.33	0.87	0.33	0.00	0.00	0.67	1.87	1.33	1.00	0.20	1.33
	Percentile.Exc 10	1.67	0.33	0.33	1.00	0.67	0.00	0.00	0.67	2.00	1.33	1.00	0.33	1.33
	Percentile.Exc 25	1.67	0.33	0.33	1.67	0.67	0.33	0.33	1.00	2.33	1.67	1.33	0.67	1.67
	Percentile.Exc 75	3.33	0.33	0.67	2.67	1.00	0.67	0.67	1.67	3.00	2.67	1.67	2.33	2.33
	Percentile.Exc 90	4.60	0.67	1.00	2.93	1.33	0.67	0.93	1.67	3.27	3.00	2.00	3.00	2.60
	Percentile.Exc 95	5.33	0.67	1.33	3.00	1.67	0.67	1.00	2.00	3.33	3.33	2.33	3.33	2.67

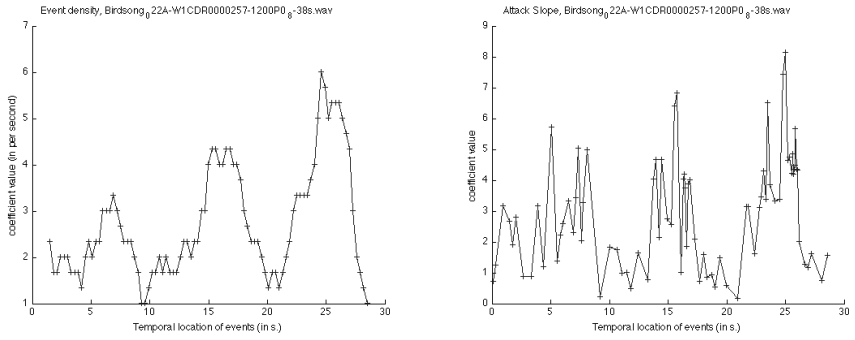


Figure 6.2.2 Variation of event density with time based on the envelope method of event detection (left) and event attack slope (right), illustrated with a birdsong recording, corresponding to the commands as below in Matlab:

o=mironsets('folder','Filter','Filterbank',0,'log','Contrast',0.03,'frame');
ed=mireventdensity(o)
o=mironsets('folder','Filter','Filterbank',0,'log','Contrast',0.03);
as=mirattackslope(o,'Contrast', 0.03)

To focus on as small number of indices as possible, only the parameters of event density of the whole duration based on both event detection methods – which may be of relative importance – are used for analysis in the next chapters.

6.2.3 Attack slope

For events detected based on the envelope method according to the maximum amplitudes, it is useful to further estimate the attack phase of each event. Attack phase is determined from the amplitude envelope for event detection by local maximum of the envelope as ending position of the attack phase and the preceding local minimum as starting position (Lartillot *et al.* 2008). Attack phase can be described by its average slope, which is equal to the ratio of the magnitude difference between beginning and ending of attack period, to the corresponding time difference (Lartillot 2011). While value of amplitude of envelope indicates the amplitude or intensity of event, attack slope indicates the increment of amplitude in a unit time. The attack slopes of events for the 13 sound recordings are computed, for which an example is shown in Figure 6.1.3 (right), using a birdsong recording.

Table 6.2.4 Statistics of attack slope over time based on different methods

AS		Birdsong	Church bells	Fountain	Machine	Music	River	Seawaves	Stream	Traffic	Voice	Wind	Big Ben	Victoria Park
Envelope	Average	2.93	2.80	0.28	1.31	0.77	0.26	0.75	0.65	1.56	1.67	0.74	1.28	1.24
o=mironsets('folder',	Median	2.81	2.57	0.22	0.76	0.65	0.24	0.58	0.47	1.35	1.39	0.53	1.11	1.12
'Filter','Filterbank',0,	STDEV	1.80	1.71	0.20	1.17	0.55	0.17	0.45	0.49	0.89	1.02	0.55	0.82	0.64
'log','Contrast',0.03);	MIN	0.18	0.34	0.05	0.22	0.17	0.04	0.35	0.12	0.26	0.35	0.16	0.19	0.25
as=mirattackslope(o,	MAX	8.14	6.01	0.83	4.02	2.23	0.64	1.92	1.80	4.35	4.23	2.24	3.46	2.46
'Contrast',0.03)	MAX-MIN	7.96	5.67	0.78	3.80	2.06	0.60	1.57	1.68	4.08	3.89	2.08	3.26	2.21
mirgetdata(as)	Percentile.Exc 5	0.55	-	0.05	0.24	0.21	-	-	0.13	0.50	0.48	0.19	0.29	0.38
	Percentile.Exc 10	0.78	0.49	0.06	0.25	0.29	0.05	0.38	0.15	0.59	0.56	0.23	0.36	0.49
	Percentile.Exc 25	1.43	1.64	0.12	0.38	0.36	0.11	0.51	0.29	0.92	0.80	0.31	0.59	0.75
	Percentile.Exc 75	4.21	4.35	0.43	1.86	0.91	0.36	0.69	1.00	2.07	2.40	1.04	1.94	1.83
	Percentile.Exc 90	5.03	5.66	0.59	3.35	1.69	0.53	1.68	1.49	2.66	3.16	1.74	2.33	2.33
	Percentile.Exc 95	6.49	-	0.80	3.56	2.14	-	-	1.75	3.88	3.73	1.95	3.12	2.38
Spectral Flux	Average	5.9	123.7	19.2	135.2	46.4	13.6	9.5	12.7	35.7	40.6	7.2	133.2	27.8
o=mironsets('folder',	Median	5.6	131.7	18.9	130.5	44.5	13.4	10.5	12.3	34.3	37.3	7.0	137.5	26.0
'SpectralFlux',	STDEV	2.1	27.6	2.1	25.7	10.4	1.3	4.9	1.8	9.4	12.7	0.9	51.6	7.3
'Contrast',0.2);	MIN	2.9	46.5	15.9	76.1	29.2	11.2	4.4	10.5	20.0	23.4	5.6	74.0	19.4
sf=get(o,'PeakVal')	MAX	12.1	151.4	27.3	238.4	85.0	18.3	18.7	24.5	67.4	79.7	11.4	281.9	60.1
for i=1:13	MAX-MIN	9.2	104.9	11.4	162.4	55.8	7.1	14.3	14.0	47.4	56.3	5.8	207.9	40.7
sf{1,i}{1,1}{1,1}	Percentile.Exc 5	3.3	-	16.7	100.6	31.9	11.8	-	10.8	23.1	26.6	6.0	74.4	20.4
end	Percentile.Exc 10	3.4	60.0	17.0	107.5	35.2	12.0	4.5	11.1	25.0	28.2	6.2	77.6	20.8
	Percentile.Exc 25	4.1	116.1	17.7	116.7	39.5	12.6	4.6	11.5	28.8	31.2	6.6	87.5	22.9
	Percentile.Exc 75	7.3	136.9	20.2	151.6	51.5	14.3	13.4	13.2	40.8	47.5	7.5	162.6	30.3
	Percentile.Exc 90	8.9	149.2	21.5	168.3	60.9	15.5	17.3	14.7	49.0	59.0	8.5	199.7	36.6
	Percentile.Exc 95	9.4	-	23.6	184.0	67.5	16.1	-	15.7	55.7	72.8	8.9	274.2	44.6

Alternatively, for the event onsets detected based on the spectral flux method, attack sharpness is directly indicated by amplitude of local maximum of the spectral flux curve, shown in Table 6.1.1. Both the methods, i.e. attack slope in amplitude envelope and spectral flux, estimate the rising amplitude in unit time, though differ in the measuring.

A number of basic, statistic indices are used to describe the results, including mean, median, standard deviation, maximum, minimum, range, and percentiles, as shown in Table 6.2.4. From the results, roughly, it can be seen that for birdsong and church bells, the average values of attack slopes of events detected by the envelope method are higher than the others, which means the event attacks of these sounds are sharper based on this measuring method. For attack sharpness calculated by spectral flux, the average values of church bells, machine, and church bells are large comparing to the others. In order to reduce the number of indices those are to be used for analysis in the next chapters, the representative indices (by checking the correlations between the indices) are remained, which are average, standard deviation, range, and 10% and 90% percentiles of both attack slope in envelope and spectral flux.

In sum, a series of indices are developed to characterize the events in sound based on the occurrence time of events or the event detection curve, including mean, median, mode, standard deviation, range, and percentiles (or part of) of each of the parameters of event interval, event density and attack slope. While additional parameters, besides periodicity, may be available, e.g. amplitude variability of event detection curve – calculated by summing the amplitude difference between successive local extrema (Lartillot *et al.* 2008), the current study focuses on the parameters discussed above.

The commands in Matlab with MIRtoolbox responding to the algorithms for the rhythm computation are

```
o=mironsets('folder','Filter','Filterbank',0,'log','Contrast',0.03);
ot=mirgetdata(o)
```

for event interval and event density based on envelope method of event detection, and

```
o=mironsets('folder','SpectralFlux','Contrast',0.2);
ot=mirgetdata(o)
```

for event interval and event density based on spectral flux method of event detection;

```
o=mironsets('folder','Filter','Filterbank',0,'log','Contrast',0.03);
as=mirattackslope(o,'Contrast',0.03);
mirgetdata(as)
```

for event attack based on envelope method, and

```
o=mironsets('folder','SpectralFlux','Contrast',0.2);
sf=get(o,'PeakVal')
```

for event attack based on spectral flux method.

6.3 Periodicity

Periodicity, in addition to the parameters discussed above that focus on general description of event detection curve, is estimated to represent the characteristics of repetition of events detected, event detection curve, or temporal variation of signal directly. To some degree, it can be seen as responding to tempo in music. Whereas rhythm in music involves hierarchy, the periodicity analysis in the current study only estimates whether events in signal exhibit periodicity, regardless the level or time scale at which it happens.

In this section, different methods for periodicity estimation are implemented, which are based on the computation of autocorrelation function of event detection curve or envelopes in filter bands, or beat spectrum from similarity matrix. By comparison of the results obtained by the different methods, the one that is to be used for periodicity estimation in analysis in the next chapters is selected.

6.3.1 Periodicity calculation based on autocorrelation of event detection curve

As a measure of the frequency of occurrence of events following an event at time zero, autocorrelation method can be used for calculation of periodicity of events, according to Brown's (1993) study that used this method for determination of musical meter from score events. Peaks in the autocorrelation function indicate the time of periodicity (a single measure in music).

In this section, the autocorrelation method is implemented with MIRtoolbox in Matlab. The autocorrelation function can be calculated from event detection curve or events already detected in Section 6.1, similar to the notated score used in Brown's study. Here, event detection curve is used. An example is given using a sound recording of church bells, shown in Figure 6.3.1. The autocorrelation functions are computed from the event detection curves of two event detection methods, i.e. methods based on overall envelope and on spectral flux. The frequency range of periodicity (the same as lowest and highest tempos in music) that is taken into consideration is above 5 bpm (5 to 6000 bpm), corresponding to repetition time of less than 12 s. The peaks in the functions indicate the most probable periodicities. With this method, the autocorrelation functions and the corresponding highest peaks of the 13 sound recordings are computed, shown in Table 6.3.1.

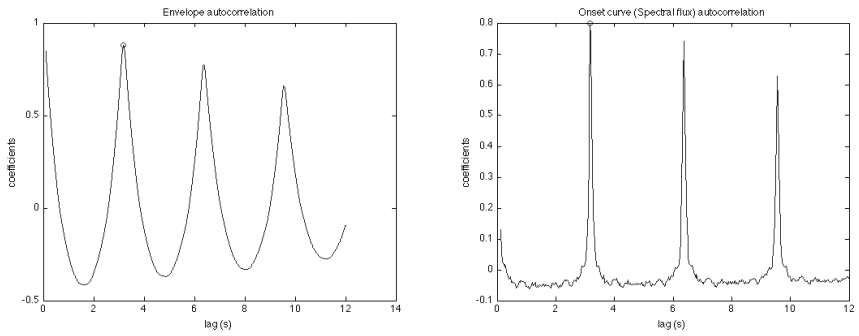


Figure 6.3.1 Periodicity calculation based on autocorrelation functions of event detection curves calculated by overall envelope (left) and by spectral flux (right), illustrated with a church bells recording, corresponding to the commands as below in Matlab:

o=mironsets('Church bell','Filter','Filterbank',0,'log','Contrast',0.03);
[t ac]=mirtempo(o,'Min',5,'Max',600,'diff',0,'Enhanced',0,'Resonance',0)
o=mironsets('Church bell,'SpectralFlux','Threshold',0.2);
[t ac]=mirtempo(o,'Min',5,'Max',600,'diff',0,'Enhanced',0,'Resonance',0)

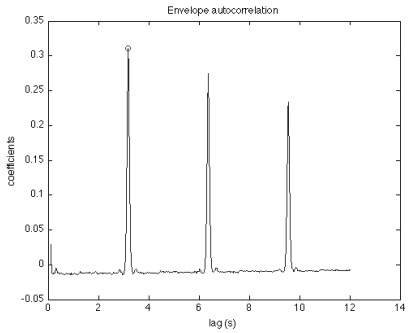


Figure 6.3.2 Periodicity calculation based on the method of envelopes in filter bands, illustrated with a church bells recording, corresponding to the commands as below in Matlab:

[t ac]=mirtempo('Church bell,'Min',5,'Max',600,'Autocor','diff',0,'Enhanced',0,'Resonance',0,
'FilterbankType','Scheirer','HalfwaveDiff','Sum','After')

6.3.2 Periodicity calculation based on envelopes in filter bands

Another method proposed by Scheirer (1998) stimulated the tempo of musical signals with strong beat by analysing separately the periodicities of signals in frequency bands and combining results at the end, rather than by stimulating on the whole signal or on the sum of filterbands. It is according to a psychoacoustic hypothesis regarding rhythmic perception that *"some sort of cross-band rhythmic integration, not simply summation across frequency bands, is performed by the auditory system"* (Scheirer 1998). This rhythmic algorithm first used a filterbank to divide the signal into a small number of

bands and extracted the amplitude envelope of each of the subbands. Then, each envelope derivative, i.e. half-wave rectified first-order difference function of the envelope, is passed on to another filterbank of tuned resonators (parallel comb filters), for one of which the resonant frequency matches the rate of periodic modulation of the envelope derivative. The frequency and phase information of the matching resonator for each of the bandpass channels are summed across to arrive at the frequency of the pulse in a rhythmic signal, i.e. the tempo or rate of the rhythm. The filterbank used in this method can alternatively be of pre-defined frequency ranges, i.e. some narrowband frequency components that are spaced apart are combined across the pre-defined ranges and the bandwidth results of onset detection and periodicity analysis are summed at later stage. This filterbank is developed to overcome the problem that frequency (e.g. pitch or harmonic) changes are easily unnoticed with envelopes of only few subbands, while individual envelopes of a large number of narrow bands are not reliable to reveal the periodicity (Goto and Muraoka 1995; Klapuri *et al.* 2006).

In this section, this method for periodicity stimulation is implemented with MIRtoolbox in Matlab. Although there are several advantages to use the approach of banks of parallel comb filters over previous autocorrelation method for detecting periodic energy modulations in a music signal with strong beat, – such as comb filtering method implicitly encodes aspects of rhythmic hierarchy and is phase preserving (Scheirer 1998), environmental sounds may generally not exhibit strong beat or beat as expected, except for certain types of sound. Since with comb filters the best matching frequency may not accurately reflect that of a signal with ambiguous or no beat, e.g. in the validation of performance of the algorithm the music samples evaluated by listeners to have no beat were not accurately beat-tracked (Scheirer 1998), autocorrelation method instead is used here for analysis of environmental sounds.

The implementation involves first a filterbank that divides signal into six bands with the borders at 0.2, 0.4, 0.8, 1.6 and 3.2 kHz according to Scheirer (1998). The amplitude envelope of each of the bands is extracted, with full-wave rectification and low-pass filtering (IIR). Then, of each envelope derivative, i.e. half-wave rectified first-order difference function of the envelope, autocorrelation is computed to examine the periodic modulation. The frequency range for the calculation of autocorrelation is the same as that in the above method, corresponding to repetition time of less than 12 s. Finally, the autocorrelation functions are summed across bands, of which the peaks of respond to the frequency of pulses in a rhythmic signal, i.e. the tempo or rate of the rhythm. An example of the summed autocorrelation function and selected peak is shown in Figure 6.3.2, illustrated with a sound recording of church bells, the same as that in Figure 6.3.1. With

this method, the periodicities of the 13 sound recordings are computed, shown in Table 6.3.1.

6.3.3 Periodicity calculation based on beat spectrum

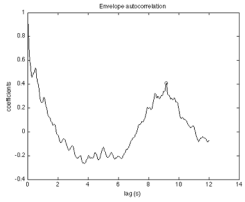
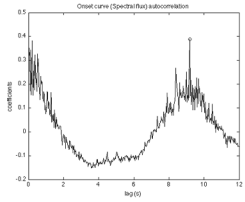
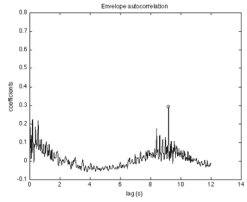
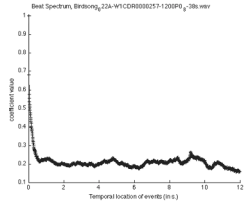
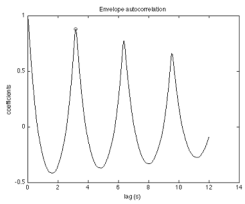
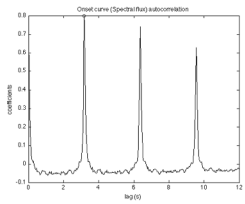
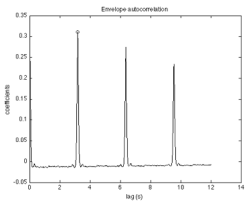
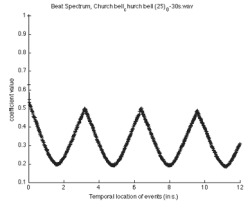
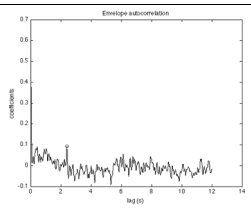
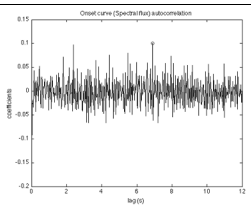
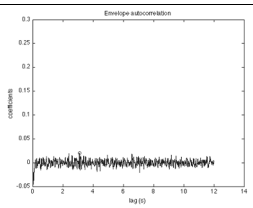
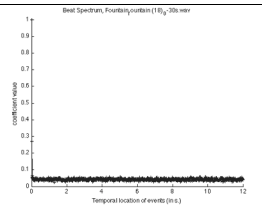
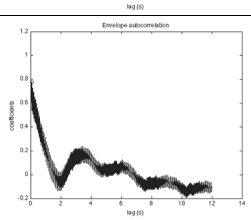
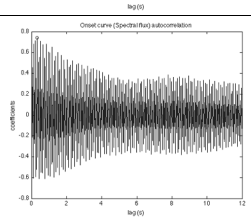
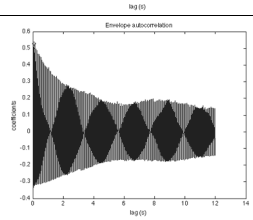
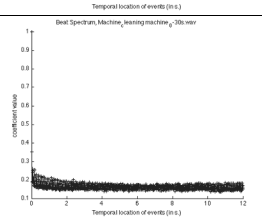
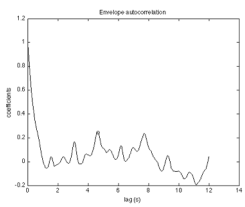
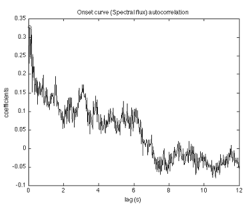
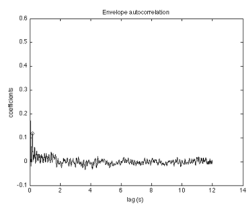
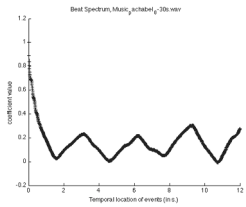
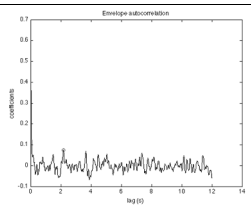
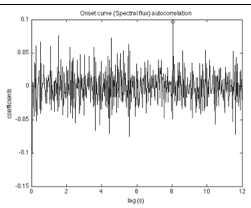
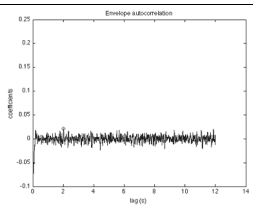
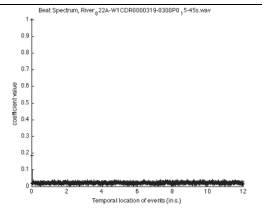
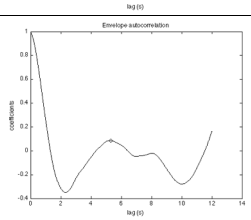
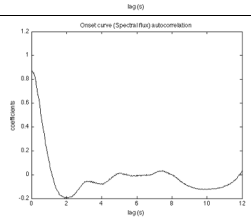
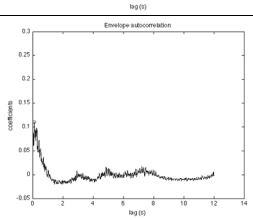
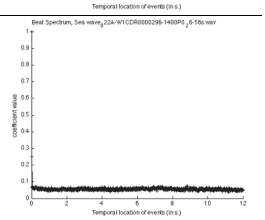
Additional method for periodicity simulation was based on similarity matrix (Foote and Cooper 2003) that described in Section 6.1.3. Essentially, similarity matrix, and the following beat spectrum, is somehow similar to the calculation of autocorrelation, except providing a visual display of similarities between instants, similar to products in autocorrelation, in a signal. However, the similarity can be calculated from multiple vectors that parameterize the instant or frame, rather than one.

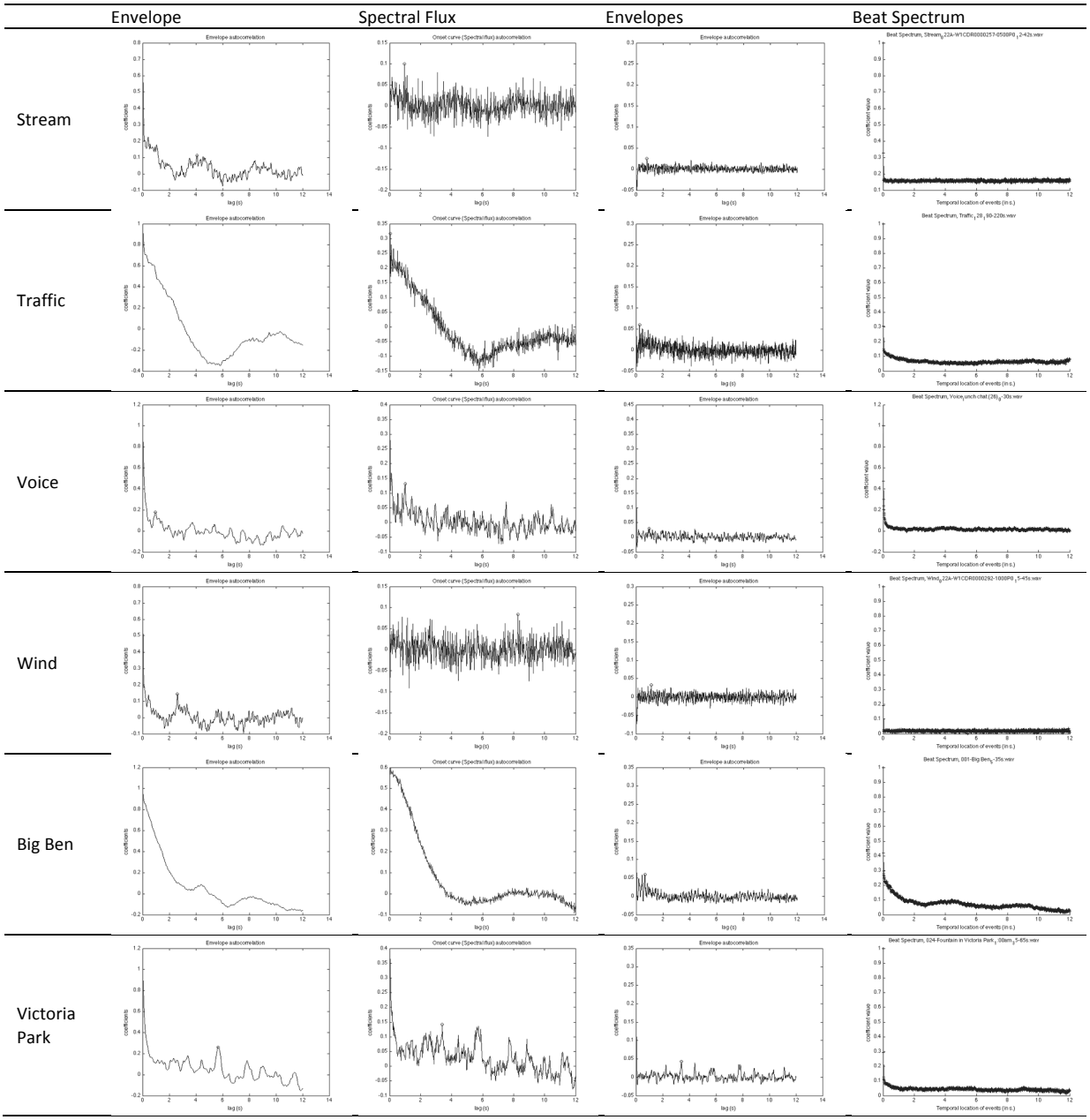
In this section, this method, named as *beat spectrum*, i.e. a measure of acoustic similarity as a function of lag time (Foote *et al.* 2002), is implemented with procedure of three main steps. First, the audio waveform is window or frame decomposed. Here, a frame length of 25ms with 10ms overlapping is used. For each frame, the short signal segment is characterized using a spectral or other features. Mel-frequency cepstral coefficients (MFCCs) are used here, computed from the spectrum that calculated by a Fourier transform (it has been tested that results calculated through spectrum and MFCCs do not differ much for the 13 samples). Then, the distance or dissimilarity between the feature vectors (MFCCs) is measured for all possible frame combinations, using cosine similarity (an alternative to Euclidean distance) (Foote and Cooper 2003). The value of dissimilarity is transformed into the value of similarity, by a simple function of that one minus value of dissimilarity. The results of inter-frame similarity between all pairs of frames are embedded into a *similarity matrix*, i.e. a two-dimensional representation which can visually show the similarities between all frames or instants in a signal. Finally, a *beat spectrum* is derived from the similarity matrix, representing the similarity in the "lag" domain where lag is the time difference between pair of frames (Foote and Cooper 2003). It is computed by either summing similarity results along the diagonal, as used here, or calculating an autocorrelation of the matrix. In the beat spectrum, the lag times of peaks correspond to repetition times of periodicity, while the amplitudes of peaks reflect relative strengths of corresponding periodicity.

The beat spectra of the 13 sound recordings are computed using the procedure, shown in Table 6.3.1. Similar to those in the methods in the last sections, the largest lag time of 12 s is taken into consideration, corresponding to 5 bpm. In beat spectrum, the peak with the maximum amplitude is selected, with the condition of parameter of "*Contrast*" of 0.05, which means the peak produce an increment larger than 0.05 of the

maximum amplitude in the spectrum without the restriction of threshold of absolute amplitude.

Table 6.3.1 Periodicity calculation of 13 sounds based on different methods

	Envelope *o=mironsets('folder','Filter','Filterbank',0,'log','Contrast',0.03);* *[t ac]=mirtempo(o,'Min',5,'Max',6000,'diff',0,'Enhanced',0,'Resonance',0)*	Spectral Flux *o=mironsets('folder','SpectralFlux','Contrast',0.2);* *[t ac]=mirtempo(o,'Min',5,'Max',6000,'diff',0,'Enhanced',0,'Resonance',0,'Contrast',0.2)*	Envelopes *[t ac]=mirtempo('folder','Min',5,'Max',6000,'Autocor','diff',0,'Enhanced',0,'Resonance',0,'FilterbankType','Scheirer','HalfwaveDiff','Sum','After','Contrast',0.1)*	Beat Spectrum *bs=mirbeatspectrum12s('folder','Dia g');* *pbs=mirpeaks(bs,'NoBegin','NoEnd','Contrast',0.05,'Total',1)*
Birdsong				
Church bells				
Fountain				
Machine				
Music				
River				
Sea waves				



6.3.4 Comparison of periodicity calculation methods

In Table 6.3.1, the periodicity results, including autocorrelation functions (or beat spectrum) and selected peak calculated by the different methods as above, of the 13 sounds are shown. The repetition times and strengths of periodicity are shown in Table 6.3.2, which respectively correspond to the lag times and amplitudes of peaks in the autocorrelation functions and beat spectrum.

From the results, it can be seen that in terms of repetition time or frequency of periodicity, for some types of sound that may exhibit clear periodicity, such as birdsong, church bells and voice, the results generally coincide through the different periodicity

calculation methods. While for the types of sound that may not show periodicity, such as fountain, river, stream, traffic and wind, the results differ from each other through the different methods – more specifically, the repetition times computed by the first three methods differ, and those by last one (beat spectrum) do not show any periodicity. In terms of strength of periodicity, the results show that for the sounds those generally have coincident results of repetition time through different methods, i.e. those may be expected to exhibit clear periodicity, the strengths of periodicity are relatively high (except for voice) by almost all the methods; contrarily, for the sounds those have different results of repetition time with different methods, the strengths of periodicity are generally low by all the methods.

Table 6.3.2 Repetition time and strength of periodicity based on different methods

		Birdsong	Churchbell	Fountain	Machine	Music	River	Seawave	Stream	Traffic	Voice	Wind	Big Ben	Victoria Park
Envelope *o=mironsets('folder','Filter','Filterbank',0,'log','Contrast',0.03);* *[t ac]=mirtempo(o,'Min',5,* *'Max',6000,'diff',0,'Enhanced',0,'Resonance',0)* *mirgetdata(ac)* *mirgetdata(t)* *pv=get(ac,'PeakVal')* *for i=1:13* *pv{1,i}{1,1}{1,1}* *end*	Repetition time	9.19	3.19	2.40	0.06	4.66	2.17	5.32	4.08	-	0.96	2.61	-	5.66
	Tempo (bpm)	6.5	18.8	25.0	928.7	12.9	27.6	11.3	14.7	-	62.4	23.0	-	10.6
	Strength	0.41	0.88	0.09	0.78	0.24	0.08	0.08	0.11	-	0.18	0.14	-	0.26
Spectral Flux *o=mironsets('folder','SpectralFlux','Contrast',0.2);* *[t ac]=mirtempo(o,'Min',5,* *'Max',6000,'diff',0,'Enhanced',0,'Resonance',0,'Contrast',0.2)* *mirgetdata(ac)* *mirgetdata(t)* *pv=get(ac,'PeakVal')* *for i= 1:13* *pv{1,i}{1,1}{1,1}* *end*	Repetition time	9.19	3.18	6.90	0.33	0.08	8.05	-	0.95	0.05	0.99	8.27	-	3.40
	Tempo (bpm)	6.5	18.8	8.7	184.3	773.5	7.5	-	63.2	1116.9	60.6	7.3	-	17.7
	Strength	0.39	0.80	0.10	0.74	0.33	0.10	-	0.10	0.32	0.13	0.08	-	0.14
Envelopes *[t ac]=mirtempo('folder',* *'Min',5,'Max',6000,'Autocor','diff',0,'Enhanced',0,'Resonance',0,'FilterbankType','Scheirer','HalfwaveDiff','Sum','After','Contrast',0.1)* *mirgetdata(ac)* *mirgetdata(t)* *pv=get(ac,'PeakVal')* *for i= 1:13* *pv{1,i}{1,1}{1,1}* *end*	Repetition time	9.19	3.19	3.10	0.07	0.16	2.01	0.11	0.80	0.26	1.00	1.14	0.65	3.41
	Tempo (bpm)	6.5	18.8	19.3	919.0	366.5	29.8	554.5	75.2	230.5	60.1	52.8	92.8	17.6
	Strength	0.29	0.31	0.02	0.53	0.12	0.02	0.11	0.02	0.06	0.03	0.03	0.06	0.04
Beat Spectrum *bs=mirbeatspectrum12s('folder','Diag');* *pbs=mirpeaks(bs,'NoBegin','NoEnd','Contrast',0.05,'Total',1)* *pp=get(pbs,'PeakPosUnit')* *for i=1:13* *pp{1,i}{1,1}{1,1}* *end* *pv=get(pbs,'PeakVal')* *for i= 1:13* *pv{1,i}{1,1}{1,1}* *end*	Repetition time	9.20	6.40	-	0.14	9.31	-	-	-	-	-	-	-	-
	Strength	0.26	0.50	-	0.25	0.30	-	-	-	-	-	-	-	-

Thus, in order to obtain relatively reliable result for both sounds with and without obvious periodicity, the method of beat spectrum is selected among the four for periodicity estimation of environmental sounds in the next chapters. Since it produces result that is coincident with those by the other methods for periodic sounds, and is showing no periodicity for sounds which have low periodicity strengths calculated by the other methods.

6.4 Periodicity Parameters

A number of descriptors can be developed for the characterization of periodicity based on statistical description of the autocorrelation function (or beat spectrum). They may be considered as a higher level dimension that contribute to subjective perception of rhythmic pulsation in sound (Lartillot *et al.* 2008), comparing to those derived from event detection curve in Section 6.2.

The evident descriptors may include the amplitude of the main peak, i.e., the global maximum of the autocorrelation function curve within a frequency range considered. It reflects the strength of periodicity. Another is the lag time of the main peak, which corresponds to repetition time of the corresponding periodicity (similar to tempo in music). In addition, the kurtosis of the main peak measures the stableness of the corresponding periodicity. A clear peak with significantly sharp slopes (high value of kurtosis) is related to a precise and stable periodicity. On the contrary, a peak displays less sharpness and gradual slopes if the periodicity fluctuates – slightly oscillates around a range of possible periodicities. The entropy of the autocorrelation function measures the peakiness, i.e. the simplicity of the function. Periodic signals tend to exhibit clearer peaks in autocorrelation functions and thus have lower entropy than nonperiodic ones. More descriptors may include harmonic relations between the peaks of autocorrelation function (Lartillot *et al.* 2008).

In the periodicity estimation of environmental sounds in the next chapters, however, in order to focus on a limited number of indices for simplification, sound is only measured by whether it exhibits periodicity based on the method of beat spectrum, regardless of repetition time (tempo) or strength of the periodicity. The commands in Matlab with MIRtoolbox corresponding to the algorithm are

```
bs=mirbeatspectrum12s('folder','Diag');
pbs=mirpeaks(bs,'NoBegin','NoEnd','Contrast',0.05,'Total',1);
pp=get(pbs,'PeakPosUnit')
```

The complete program is available in the accompanying CD-ROM.

6.5 Conclusions and Discussions

With systematic exploration of rhythm algorithms proposed in literature, especially in music information extraction, a number of rhythm models are implemented with Matlab program. By examining their performances with 13 types of common environmental sound in soundscapes, both the methods based on overall envelope and on spectral flux are selected for sound event detection, and the method of beat spectrum is selected for periodicity estimation of environmental sounds in this study. The parameters of the models have been adjusted for environmental sounds.

A number of parameters are derived from the event detection and periodicity estimation to describe the rhythm feature of environmental sounds, which are event interval, event density, event attack slope and spectral flux, as well as periodicity. A series of statistic indices are thus developed to describe the rhythm parameters with time, including mean, median, mode, standard deviation, range and percentiles (or part) of each of the parameters, for the further rhythm analysis in this study.

While from the systematic reviews of definitions and algorithms of the psychoacoustic and music parameters in these two chapters (Chapters 5 and 6) and in Chapter 2, it is expected that there may be certain correlations between the parameters of pitch strength and tonality, and between rhythm and fluctuation, the correlations between the pitch and rhythm indices developed in these two chapters and the previous psychoacoustic (loudness and timbre) indices analysed in Chapter 4 are examined in the following chapter. Also, these pitch and rhythm indices are further refined based on the larger sound sample in this study by the examination of their correlations and principal components, in terms of which the characteristics of different types of sound are explored. Moreover, the contribution of the pitch and rhythm parameters to the automatic identification of environmental sound type is examined in the following chapter.

Chapter 7
Characteristics of natural and urban sounds in soundscapes in terms of pitch and rhythm features

In this chapter, the characteristics of natural and urban sounds in soundscapes are analysed in terms of music features, including pitch and rhythm indices as discussed and developed in Chapters 5 and 6. While the pitch and rhythm features are on a psychoacoustic basis, they had their application mainly in music and speech rather than environmental sound; for convenience, they are termed as music features in this chapter, in order to be distinguished from the psychoacoustic parameters (loudness and timbre in general) that have been analysed with in Chapter 4.

In this chapter, the correlations between the developed pitch and rhythm indices and the loudness and timbre indices used in Chapter 4 are firstly examined in Section 7.1. Then, the principal components of the pitch and rhythm indices are analysed respectively in Sections 7.2 and 7.3, as well as the differences among the types of sound in terms of the principal components. Finally, in Section 7.4, the correlations between and the principal components of all the pitch and rhythm indices are discussed.

7.1 Correlations Between the Pitch and Rhythm Indices and the Loudness and Timbre Indices

The indices used for analysis in this chapter include the pitch and rhythm indices that developed in the last two chapters. For pitch feature, they are values of the best four average pitches for the whole duration (PV1, PV2, PV3, PV4) and their corresponding pitch strengths or amplitudes (PA1, PA2, PA3, PA4), the percentage of audible pitches over time (PN), and a number of statistic indices of pitch values (PV) and pitch strengths (PA) over time, which include average (PV AVE, PA AVE), median (PV Median, PA Median), mode (PV Mode), standard deviation (PV STDEV, PV STDEVA, PA STDEV, PA STDEVA), range (PV Range, PA Range), and percentiles (PV Percentile5, PV Percentile25, PV Percentile75, PV Percentile95, PA Percentile5, PA Percentile95).

For rhythm feature, a number of parameters are derived, which are event interval or interonset interval (IOI), event density (ED), and attack slope (AS) or spectral flux (SF), based on the event detection methods of both overall envelope and spectral flux

(indicated by (E) and (SF) respectively), as well as periodicity based on the method of beat spectrum (BS). A series of statistic indices are included to describe the parameters with time, which are average (IOI(E) AVE, IOI(SF) AVE, AS AVE, SF AVE), median (IOI(E) Median, IOI(SF) Median), mode (IOI(E) Mode, IOI(SF) Mode), standard deviation (IOI(E) STDEV, IOI(SF) STDEV, AS STDEV, SF STDEV), range (IOI(E) Range, IOI(SF) Range) and percentiles (IOI(E) Percentile10, IOI(E) Percentile90, IOI(SF) Percentile10, IOI(SF) Percentile90, AS Percentile10, AS Percentile90, SF Percentile10, SF Percentile90). The pitch and rhythm indices are also displayed in Table 7.2.1 and Table 7.2.2 respectively. The analyses in this chapter are based on the data of the first segment of each of the 102 recording.

In order to check if any of these pitch and rhythm indices represent the same or similar variance with the psychoacoustic indices that have been used in Chapter 4, in other words, if there are any indices which are highly correlated, the correlations between the pitch and rhythm indices and the loudness and timbre indices are examined in this section. Table 7.2.1 shows the correlations of the pitch indices with the psychoacoustic ones, where ** 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. These symbols remain the same meanings in the following tables in this chapter. From Table 7.2.1, it can be seen that the correlation coefficients generally are not very high, with the highest coefficients of 0.6 to 0.8. These relatively high correlations are between the pitch indices and S STD, S MAX, Ton MAX, Fls AVE, Fls STD, and Fls MAX, while correlations between the pitch indices and the rest psychoacoustic indices are generally below 0.6, except for that between PV Range and S AVE (of 0.61). In terms of the relatively high correlations, specifically, almost all the PV and PA statistic indices over time have high correlations (above 0.6) with S STD and S MAX, besides PV1 and PA1 have high correlations with S STD, and PV1 and PV2 have high correlations with S MAX. Parts of the PA indices, which are PA2, PA3, PA AVE, PA Median, PA STDEVA and PA Range, have high correlations with Ton MAX. Although it would be expected that pitch strength would have certain correlation with tonality as discussed in Chapter 5, the results show that the correlation is not high. For fluctuation, the PV indices including PV AVE, PV STDEV, PV STDEVA, PV Percentile75 and PV Percentile95 have high correlations with Fls AVE, Fls STD, and Fls MAX; for PA indices, PA AVE, PA Median, PA STDEVA, PA Range and PA Percentile5 have high correlations with Fls AVE, and PA STDEVA has high correlations with Fls STD and Fls MAX. In general, the above results reveal that there are certain correlations between the pitch (both value and strength) and the variation and maximum of sharpness, between pitch strength and maximum of tonality, and between both value

and strength of pitch and fluctuation strength, however the correlations are not very high. These correlations 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, e.g., some samples show a number of certain characteristics.

The correlations of rhythm indices with psychoacoustic indices are shown in Table 7.2.2. Similar to those of pitch, it can be seen the correlation coefficients are not very high, generally below 0.8. For AS indices, AS AVE, AS Percentile10 and AS Percentile90 have relatively high correlations (of coefficient of about 0.6 to 0.7) with S STD and Fls AVE; AS AVE and AS Percentile90 have relatively high correlations with S MAX, Fls STD and Fls MAX. Almost all the SF indices have high correlations (between 0.6 and 0.9) with L AVE, L MAX, L MIN, N AVE, N STD and N MAX. Also, IOI(SF) STDEV, IOI(SF) Range and ED(SF) have high correlations (between 0.6 and 0.7) with L STD; BS has high correlations (about 0.6) with S STD and Ton MAX. These results reveal certain correlations between attack slope and fluctuation (both the STD and Fls indices), between spectral flux and SPL and loudness, and between periodicity and variation of sharpness and maximum of tonality; however, the correlations are generally not very high, except for those between spectral flux and maxima of SPL and loudness, which reach over 0.8. The high correlations between spectral flux and maxima of SPL and loudness may somehow be expected, as spectral flux is calculated from the differences of SPLs with time at different spectral frequencies.

Based on the results above, it can be concluded that generally the pitch and rhythm indices developed in the last two chapters provide additional variance to the psychoacoustic indices that have been used for analysis in Chapter 4, since generally the correlations between the indices are not high, although there are certain correlations between, e.g., pitch and sharpness, pitch strength and tonality, attack slope and fluctuation, and spectral flux and SPL or loudness.

7.2 Characteristics of Natural and Urban Sounds in Soundscapes in Terms of Pitch Features

The characteristics of natural and urban sounds in soundscapes are analysed in terms of the pitch indices in this section. Firstly, the correlations of the pitch indices are examined to see whether some indices are highly correlated, if so, some of them can be removed from the index set. Then, the principal components of the pitch indices are explored, based on which the characteristics of natural and urban sounds are analysed subsequently. Finally, the recordings are automatically identified with discriminant function analysis.

Table 7.2.1 Pearson's correlations between pitch and psychoacoustic indices

	L AVE	L STDEV	N AVE	N STDEV	S AVE	S STDEV	Ton AVE	Ton STDEV	R AVE	R STDEV	Fls AVE	Fls STDEV	L MAX	L MIN	N MAX	N MIN	S MAX	S MIN	Ton MAX	Ton MIN	R MAX	R MIN	Fls MAX	Fls MIN
PV1	-.352**	.493**	-.429**	0.040	.395**	**.620****	-0.035	-0.003	-.442**	0.112	.419**	.483**	-0.051	-.391**	-0.177	-.455**	**.654****	-0.017	0.168	-0.086	-0.204	-.391**	.508**	0.205
PV2	-0.238	.386**	-.335**	0.083	.411**	.520**	-0.043	-0.076	-.324*	0.064	.392**	.441**	0.051	-.293*	-0.070	-.365**	**.638****	0.056	0.114	.c	-0.192	-.298*	.489**	0.244
PV3	-.406**	.459**	-.467**	0.016	.327*	.480**	-0.074	-0.109	-.514**	0.059	0.200	.294*	-0.143	-.465**	-0.264	-.491**	.491**	-0.068	0.079	.c	-.304*	-.478**	.330*	-0.029
PV4	-.324*	0.230	-.372*	-0.036	.319*	.290*	0.021	0.025	-.358*	-0.076	.321*	0.166	-0.193	-.316*	-0.231	-.369*	.352*	0.044	0.102	.c	-.298*	-.305*	0.273	.291*
PA1	-.233*	.530**	-.365**	.292*	0.156	**.643****	.382**	.480**	-.534**	0.165	.517**	.533**	0.127	-.303**	-0.033	-.460**	.539**	-.262*	.587**	0.090	-0.220	-.508**	.559**	.339**
PA2	-0.188	.469**	-.335**	.391**	0.003	.503**	.552**	.580**	-.506**	0.128	.426**	.439**	0.146	-.275*	0.037	-.446**	.418**	-.348**	**.646****	.c	-0.214	-.501**	.474**	.282*
PA3	-0.158	.557**	-.295*	.481**	0.039	.492**	.563**	.580**	-.531**	0.071	.367**	.389**	0.214	-0.274	0.123	-.413**	.400**	-.322*	**.627****	.c	-0.271	-.506**	.415**	0.178
PA4	-0.154	.521**	-0.269	.471**	0.064	.470**	.546**	.555**	-.531**	0.040	.354*	.362*	0.176	-0.248	0.103	-.387**	.385**	-0.285	.570**	.c	-.293*	-.498**	.378**	0.163
PN	.274**	-.361**	.397**	-0.002	-0.192	-.433**	-0.134	-.241*	.488**	-0.075	-.408**	-.393**	0.019	.367**	0.185	.459**	-.422**	0.149	-.431**	0.072	.217*	.470**	-.424**	-.268**
PV AVE	-.386**	.501**	-.425**	0.006	.578**	**.759****	0.051	0.100	-.422**	0.194	**.623****	**.602****	-0.072	-.460**	-0.161	-.466**	**.792****	0.081	.292**	-0.070	-0.144	-.437**	**.641****	.431**
PV Median	-.384**	.481**	-.419**	-0.008	.560**	**.715****	0.062	0.106	-.425**	0.175	.574**	.544**	-0.080	-.451**	-0.163	-.457**	**.766****	0.094	.294**	-0.061	-0.154	-.435**	.594**	.414**
PV Mode	-.247*	.429**	-.304**	0.036	.591**	**.611****	0.051	0.079	-.300**	0.190	.571**	.446**	-0.018	-.359**	-0.118	-.371**	**.683****	0.166	.252*	-0.044	-0.089	-.349**	.516**	.507**
PV STDEV	-.360**	.429**	-.391**	0.033	.545**	**.771****	0.028	0.101	-.358**	0.191	**.620****	**.624****	-0.059	-.409**	-0.107	-.424**	**.778****	0.059	.290**	-0.098	-0.089	-.370**	**.661****	.444**
PV STDEVA	-.344**	.453**	-.362**	0.048	.565**	**.740****	0.017	0.060	-.329**	.234*	**.615****	**.642****	-0.041	-.388**	-0.085	-.396**	**.769****	0.111	.213*	-0.087	-0.056	-.365**	**.677****	.405**
PV Range	-.387**	.358**	-.331**	0.014	**.612****	**.646****	-0.062	0.018	-.195*	.223*	.533**	.488**	-0.152	-.418**	-0.103	-.337**	**.715****	.236*	0.116	-0.146	0.016	-.255**	.518**	.429**
PV Percentile5	-.330**	.457**	-.370**	0.005	.568**	**.701****	0.074	0.129	-.387**	0.192	.597**	.471**	-0.091	-.441**	-0.181	-.429**	**.696****	0.076	.291**	-0.046	-0.137	-.419**	.513**	.496**
PV Percentile25	-.356**	.508**	-.391**	0.013	.516**	**.691****	0.060	0.097	-.404**	.218*	.555**	.524**	-0.058	-.439**	-0.170	-.437**	**.714****	0.070	.249*	-0.051	-0.135	-.432**	.552**	.372**
PV Percentile75	-.378**	.482**	-.423**	0.000	.562**	**.751****	0.047	0.097	-.416**	0.170	**.632****	**.630****	-0.064	-.446**	-0.154	-.459**	**.783****	0.061	.296**	-0.071	-0.147	-.423**	**.665****	.418**
PV Percentile95	-.369**	.472**	-.409**	0.034	.568**	**.817****	0.032	0.102	-.391**	.207*	**.625****	**.611****	-0.057	-.435**	-0.129	-.450**	**.808****	0.057	.301**	-0.091	-0.112	-.401**	**.649****	.463**
PA AVE	-.236*	.461**	-.304**	0.110	.419**	**.690****	.338**	.449**	-.402**	0.166	**.637****	.520**	0.030	-.373**	-0.062	-.393**	**.652****	-0.086	**.633****	0.096	-0.127	-.432**	.573**	.546**
PA Median	-.244*	.480**	-.314**	0.119	.433**	**.711****	.357**	.458**	-.420**	0.172	**.655****	.536**	0.040	-.382**	-0.056	-.406**	**.674****	-0.087	**.651****	0.120	-0.137	-.450**	.593**	.564**
PA STDEV	-0.194	.381**	-.261**	0.082	.342**	.575**	.274**	.402**	-.340**	0.122	.507**	.403**	0.003	-.324**	-0.077	-.340**	.534**	-0.089	.550**	0.034	-0.114	-.360**	.446**	.437**
PA STDEVA	-.233*	.487**	-.316**	0.171	.405**	**.739****	.426**	.547**	-.438**	.278**	**.769****	**.705****	0.082	-.361**	0.011	-.433**	**.716****	-0.118	**.708****	0.095	-0.044	-.516**	**.766****	.598**
PA Range	-.233*	.443**	-.301**	0.098	.423**	**.661****	.298**	.410**	-.379**	0.168	**.636****	.509**	0.016	-.374**	-0.065	-.397**	**.637****	-0.079	**.601****	0.074	-0.108	-.417**	.564**	.560**
PA Percentile5	-.240*	.445**	-.287**	0.085	.458**	**.677****	.278**	.388**	-.352**	0.163	**.633****	.496**	0.018	-.370**	-0.063	-.357**	**.650****	-0.037	.578**	0.087	-0.109	-.384**	.545**	.567**
PA Percentile95	-.221*	.408**	-.283**	0.095	.337**	**.615****	.324**	.439**	-.373**	0.151	.559**	.461**	0.011	-.337**	-0.069	-.363**	.567**	-0.107	.598**	0.069	-0.112	-.399**	.508**	.473**

Table 7.2.2 Pearson's correlations between rhythm and psychoacoustic indices

	L AVE	L STDEV	N AVE	N STDEV	S AVE	S STDEV	Ton AVE	Ton STDEV	R AVE	R STDEV	Fls AVE	Fls STDEV	L MAX	L MIN	N MAX	N MIN	S MAX	S MIN	Ton MAX	Ton MIN	R MAX	R MIN	Fls MAX	Fls MIN
IOI(E) AVE	.247*	-0.034	.354**	.313**	-.211*	-.249*	-0.100	-0.143	.281**	-0.007	-.380**	-.347**	0.165	.247*	.261**	.323**	-.318**	-0.005	-.257**	-0.003	0.089	.250*	-.367**	-.312**
IOI(E) Median	.221*	0.001	.337**	.341**	-0.194	-.231*	-0.088	-0.117	.262**	0.002	-.367**	-.337**	0.159	.213*	.272**	.295**	-.306**	-0.002	-.247*	0.000	0.069	.213*	-.358**	-.303**
IOI(E) Mode	0.181	-0.073	.267**	.201*	-0.132	-.235*	0.005	0.001	.204*	-0.063	-.293**	-.272**	0.082	0.183	0.192	.234*	-.263**	0.041	-0.150	0.019	0.017	0.166	-.291**	-.235*
IOI(E) STDEV	.256**	-0.045	.345**	.260**	-.216*	-.258**	-0.112	-0.165	.275**	-0.018	-.374**	-.331**	0.156	.252*	.224*	.312**	-.314**	-0.013	-.263**	-0.030	0.098	.255**	-.350**	-.318**
IOI(E) Range	.231*	-0.060	.324**	.217*	-.249*	-.289**	-0.112	-0.173	.256**	-0.032	-.388**	-.335**	0.121	.235*	.200*	.289**	-.345**	-0.038	-.268**	-0.037	0.092	.239*	-.353**	-.334**
IOI(E) Percentile10	0.178	-0.166	.303**	0.096	-.262**	-.365**	-0.007	-0.063	.240*	-0.116	-.392**	-.380**	0.055	.213*	0.177	.326**	-.418**	-0.004	-0.176	0.100	0.020	.231*	-.401**	-.279**
IOI(E) Percentile90	0.156	-0.070	.253*	0.140	-.291**	-.352**	-0.089	-0.147	0.192	-0.103	-.444**	-.393**	0.027	0.166	0.116	.231*	-.430**	-0.058	-.269**	-0.010	-0.013	0.180	-.415**	-.375**
ED(E)	-.292**	.213*	-.443**	-0.189	.312**	.454**	0.040	0.146	-.376**	0.135	.522**	.446**	-0.128	-.336**	-.285**	-.429**	.471**	-0.051	.307**	-0.065	-0.096	-.363**	.471**	.415**
AS AVE	-.310**	.545**	-.426**	0.107	.391**	**.698****	.203*	.369**	-.419**	.329**	**.706****	**.656****	0.031	-.452**	-0.104	-.520**	**.661****	-0.126	.504**	-0.074	-0.020	-.505**	**.677****	.505**
AS STDEV	-0.086	.255**	-0.136	0.117	0.184	.257**	0.025	0.085	-0.095	0.145	.258**	.235*	0.095	-0.157	0.022	-.214*	.249*	-0.023	0.145	-0.050	0.024	-0.156	.243*	0.177
AS Percentile10	-.275**	.452**	-.415**	0.051	.356**	**.600****	0.136	.300**	-.408**	.237*	**.615****	.489**	-0.006	-.432**	-0.164	-.484**	.559**	-0.118	.455**	-0.053	-0.073	-.462**	.528**	.534**
AS Percentile90	-.278**	.504**	-.374**	0.133	.321**	**.644****	.201*	.351**	-.389**	.310**	**.631****	**.621****	0.046	-.384**	-0.069	-.449**	**.617****	-0.117	.455**	-0.073	-0.002	-.456**	**.636****	.413**
IOI(SF) AVE	-0.092	.380**	-0.072	.370**	-0.063	0.134	0.067	0.140	-0.039	.255**	0.082	0.111	.244*	-0.177	.269**	-.201*	0.135	-0.150	0.138	-0.056	0.107	-0.154	0.135	0.051
IOI(SF) Median	0.026	0.171	0.030	.287**	-0.045	0.058	0.097	0.166	0.060	0.146	0.069	0.047	0.194	-0.043	.258**	-0.076	0.074	-0.079	0.119	-0.034	0.114	-0.016	0.062	0.087
IOI(SF) Mode	0.039	0.175	0.057	.288**	-0.002	0.048	.284**	.424**	0.008	0.105	0.130	0.136	0.164	0.013	.228*	0.016	0.055	-0.007	.262**	-0.019	0.073	-0.053	0.108	0.036
IOI(SF) STDEV	-.202*	**.621****	-0.170	.465**	-0.017	.271**	-0.005	0.047	-0.157	.396**	0.080	.209*	.258**	-.304**	.231*	-.318**	.208*	-.199*	0.126	-0.072	0.088	-.297**	.224*	-0.067
IOI(SF) Range	-.236*	**.708****	-.215*	.502**	0.023	.362**	0.013	0.090	-.221*	.451**	0.152	.274**	.228*	-.351**	0.188	-.364**	.262**	-.232*	.212*	-0.082	0.071	-.362**	.267**	-0.047
IOI(SF) Percentile10	-0.028	0.144	0.004	.222*	-0.062	-0.010	0.182	.296**	0.017	0.164	0.088	0.087	0.088	-0.083	0.174	-0.045	-0.010	-0.043	0.146	-0.030	0.137	-0.053	0.066	0.034
IOI(SF) Percentile90	-.206*	.421**	-0.167	.240*	-0.092	0.118	0.028	0.071	-0.159	.199*	0.009	0.065	0.133	-.272**	0.061	-.251*	0.060	-0.178	0.059	-0.060	-0.004	-.256*	0.070	-0.074
ED(SF)	.238*	**-.686****	.211*	-.522**	-0.066	-.411**	-0.116	-.215*	0.171	-.471**	-.297**	-.346**	-.240*	.384**	-.244*	.388**	-.337**	.199*	-.264**	0.080	-0.159	.351**	-.357**	-0.157
SF AVE	**.650****	.272**	.597**	**.735****	.237*	.262**	.320**	.354**	.463**	.422**	.198*	.242*	**.874****	.475**	**.818****	.367**	.358**	0.135	.337**	0.047	.534**	.279**	.257**	0.160
SF STDEV	.377**	.358**	.279**	**.607****	0.161	.311**	.213*	.266**	.248*	.492**	.287**	.336**	**.709****	.213*	**.621****	0.071	.381**	0.009	.318**	-0.005	.461**	0.064	.366**	.228*
SF Percentile10	**.773****	0.100	**.750****	**.651****	.236*	0.130	.303**	.288**	.587**	.321**	0.086	0.114	**.880****	**.618****	**.845****	.544**	.261**	.217*	.258*	0.079	.559**	.425**	0.123	0.096
SF Percentile90	.558**	.295**	.478**	**.661****	0.196	.269**	.297**	.344**	.373**	.480**	.249*	.287**	**.812****	.383**	**.712****	.270**	.341**	0.083	.346**	0.031	.527**	0.183	.293**	0.189
BS	-0.065	.467**	-0.088	.262**	.272**	**.613****	.549**	.579**	-.290**	.254**	.479**	.510**	0.165	-.195*	0.071	-.224*	.474**	-0.183	**.602****	0.190	-0.016	-.379**	.508**	.266**

Table 7.2.3 Pearson's correlations of pitch indices

	PV1	PV2	PV3	PV4	PA1	PA2	PA3	PA4	PN	PV AVE	PV Median	PV Mode	PV STDEV	PV STDEVA	PV Range	PV Percentile5	PV Percentile25	PV Percentile75	PV Percentile95	PA AVE	PA Median	PA STDEV	PA STDEVA	PA Range	PA Percentile5	PA Percentile95
PV1	1																									
PV2	.625**	1																								
PV3	.720**	.345*	1																							
PV4	.380**	.404**	0.281	1																						
PA1	.729**	.492**	.449**	.385**	1																					
PA2	.479**	.448**	0.258	.329*	**.897****	1																				
PA3	.480**	.404**	.338*	.385**	**.847****	**.971****	1																			
PA4	.445**	.349*	.316*	.431**	**.823****	**.934****	**.976****	1																		
PN	-.649**	-.457**	-.561**	-.392**	-.701**	-.580**	-.493**	-.504**	1																	
PV AVE	**.834****	.720**	.697**	.554**	.654**	.491**	.506**	.494**	-.588**	1																
PV Median	**.817****	.701**	.711**	.549**	.626**	.475**	.530**	.518**	-.589**	**.986****	1															
PV Mode	.516**	.441**	.553**	.475**	.345**	0.210	0.185	0.213	-.456**	**.818****	**.817****	1														
PV STDEV	**.841****	.767**	.607**	.465**	.693**	.532**	.470**	.429**	-.511**	**.895****	**.846****	.596**	1													
PV STDEVA	.777**	.721**	.568**	.432**	.578**	.414**	.395**	.365*	-.357**	**.899****	**.865****	.699**	**.908****	1												
PV Range	.617**	.622**	.492**	.342*	.424**	.300*	0.179	0.158	-.238*	.721**	.681**	.508**	**.806****	.769**	1											
PV Percentile5	.627**	.424**	.620**	.511**	.464**	.309*	.320*	.329*	-.557**	**.873****	**.863****	**.892****	.665**	.697**	.569**	1										
PV Percentile25	.706**	.561**	.619**	.477**	.566**	.408**	.437**	.450**	-.548**	**.935****	**.936****	**.842****	.706**	.777**	.598**	**.899****	1									
PV Percentile75	**.857****	.756**	.667**	.554**	.680**	.515**	.509**	.489**	-.588**	**.985****	**.954****	.766**	**.937****	**.913****	.727**	**.814****	**.874****	1								
PV Percentile95	**.853****	.738**	.687**	.476**	.697**	.530**	.488**	.452**	-.543**	**.951****	**.912****	.699**	**.970****	**.911****	.783**	.793**	**.823****	**.958****	1							
PA AVE	.559**	.374**	.436**	.373**	.638**	.556**	.545**	.524**	-.713**	.709**	.696**	.575**	.640**	.451**	.470**	.722**	.655**	.700**	.688**	1						
PA Median	.576**	.409**	.446**	.413**	.668**	.591**	.579**	.559**	-.714**	.741**	.728**	.613**	.661**	.497**	.482**	.748**	.691**	.730**	.713**	**.994****	1					
PA STDEV	.461**	.264*	.365**	0.266	.521**	.447**	.437**	.418**	-.679**	.558**	.547**	.417**	.520**	.272**	.376**	.582**	.496**	.556**	.552**	**.954****	**.920****	1				
PA STDEVA	.609**	.452**	.495**	.337*	.707**	.636**	.604**	.575**	-.677**	.760**	.753**	.638**	.701**	.660**	.517**	.709**	.687**	.755**	.727**	**.846****	.863**	.722**	1			
PA Range	.537**	.344**	.415**	.369*	.567**	.475**	.468**	.445**	-.717**	.678**	.668**	.577**	.601**	.423**	.448**	.719**	.626**	.667**	.647**	**.984****	**.975****	**.946****	**.845****	1		
PA Percentile5	.531**	.333**	.416**	.376**	.567**	.469**	.482**	.457**	-.633**	.701**	.687**	.597**	.617**	.446**	.489**	.747**	.659**	.683**	.675**	**.980****	**.976****	**.919****	**.804****	**.973****	1	
PA Percentile95	.526**	.298*	.417**	0.271	.600**	.501**	.479**	.463**	-.693**	.616**	.609**	.458**	.580**	.352**	.418**	.612**	.542**	.613**	.611**	**.956****	**.923****	**.976****	.781**	**.934****	**.914****	1

7.2.1 Correlations between the pitch indices

The correlations between the pitch indices are examined based on the 102 samples, shown in Table 7.2.3, where the correlation coefficients which are higher than 0.8 are highlighted with bold numbers. It shows that a number of the indices are highly correlated, e.g., the correlations are high (above 0.8) among the pitch strengths or amplitudes of the best four average pitches for the whole duration (PA1, PA2, PA3, PA4), among the indices of pitch values over time (PV), and among the indices of pitch strengths over time (PA).

Among the highly correlated indices, the ones that have particularly high correlations with some others, higher than 0.97, are removed from the index set. These removed indices are PV Median, PV Percentile75, PV Percentile95, PA Median, PA Range, PA Percentile5 and PA Percentile95 (greyed in Table 7.2.3). Their variance can be represented by the remaining indices, among which PV Median and PV Percentile75 have high correlations with PV AVE; PV Percentile95 has high correlations with PV STDEV; PA Median, PA Range and PA Percentile5 have high correlations with PA AVE; and PA Percentile95 has high correlations with PA STDEV. Although PA3 also has high correlations (higher than 0.97) with PA2 and PA4, it is remained since it forms the paired indices with PV3.

7.2.2 Principal components of the pitch indices

Since there are certain correlations between the pitch indices as discussed in Section 7.2.1, principal component analysis (PCA) is implemented with software of SPSS Statistics 20 to reduce the dimensionality of the dataset. Based on the results of the remaining pitch indices for the 102 recordings, the PCA is conducted on the correlation matrix of the 19 indices. Before implementation of the PCA, Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is taken, showing a result of 0.75. It generally indicates the adequacy of the sample size and the availability of the analysis, as a value of 0.6 is a suggested minimum.

From the 19 indices or variables, 19 components are extracted. The variances that explained by the components are shown in Table 7.2.4. It can be seen that the eigenvalues of first four components are greater than one. The first component explains 54.4% of the total variance, while the second, third and fourth explains 14.4%, 9.7% and 5.6% respectively. Table 7.2.5 shows the correlations between the first four components and indices, from which it can be seen that Component 1 has high correlations (above 0.6)

with almost all the indices, except for PV4, PV Range and PA STDEV (of below but about 0.6); Component 2 has high correlations with PA2 and PA3; Components 3 and 4 do not have any high correlation with any of the indices. These results can also be seen on the component loading plots, shown in Figure 7.2.1, where the first three components are displayed. From the plots, it can be seen that the indices are generally clustered in three groups, i.e. the index of percentage of audible pitches over time (PN), the pitch value (PV) indices, and the pitch strength (PA) indices. In the group of PA indices, PA2, PA3 and PA4 are close to each other, which have high correlations among as shown in Table 7.2.3. Component 1 mainly distinguishes PN from PV and PA indices; Component 2 mainly distinguishes between the PV indices and PA indices; and Components 3 mainly distinguishes PA AVE and PA STDEV from the rest indices.

Table 7.2.5 also shows the proportion of each index's variance that can be explained by the retained principal components. When the first four or three components are retained, the proportions of all the indices' variance that can be explained are generally high expect for PV4, of which is below 0.5, which means that generally all these indices (expect for PV4) are well represented by the principal components. When two components are retained, PV2, PV4 and PA STDEV, which have relatively low proportion values of below 0.5, are not well represented. Thus, the first three components are generally necessary to represent the original indices, which together account for 78.5% of the total variance.

Table 7.2.4 Total variance explained by the components based on pitch indices

Component	Eigenvalues	% of Variance	Cumulative %
1	**10.333**	**54.386**	**54.386**
2	**2.742**	**14.434**	**68.820**
3	**1.836**	**9.665**	**78.485**
4	**1.069**	**5.626**	**84.111**
5	0.805	4.239	88.350
6	0.565	2.975	91.325
7	0.391	2.057	93.382
8	0.346	1.823	95.205
9	0.304	1.598	96.803
10	0.181	0.953	97.756
11	0.132	0.695	98.451
12	0.089	0.469	98.920
13	0.076	0.399	99.319
14	0.067	0.354	99.673
15	0.028	0.147	99.821
16	0.014	0.074	99.895
17	0.011	0.056	99.951
18	0.006	0.034	99.985
19	0.003	0.015	100.000

Table 7.2.5 Component matrix and communalities for pitch indices

	Component matrix				Communalities		
	Component 1	Component 2	Component 3	Component 4	Extraction of 2 components	Extraction of 3 components	Extraction of 4 components
PV1	**0.807**	0.165	0.200	-0.268	.678	.718	.790
PV2	**0.628**	0.250	0.396	-0.158	**.457**	.614	.639
PV3	**0.692**	0.259	-0.102	-0.129	.546	.556	.573
PV4	0.580	0.074	0.061	0.373	**.342**	**.346**	**.485**
PA1	**0.785**	-0.422	0.283	-0.024	.794	.874	.875
PA2	**0.653**	**-0.664**	0.290	0.128	.867	.951	.968
PA3	**0.684**	**-0.631**	0.299	0.117	.866	.956	.970
PA4	**0.691**	-0.589	0.280	0.178	.825	.904	.935
PN	**-0.826**	0.142	0.387	0.170	.702	.852	.880
PV AVE	**0.939**	0.286	0.018	0.075	.963	.963	.969
PV Mode	**0.665**	0.418	-0.296	0.446	.618	.705	.904
PV STDEV	**0.835**	0.260	0.243	-0.301	.765	.825	.915
PV STDEVA	**0.744**	0.437	0.398	0.050	.745	.904	.906
PV Range	0.591	0.513	0.193	-0.339	.613	.650	.764
PV Percentile5	**0.785**	0.281	-0.382	0.291	.696	.841	.926
PV Percentile25	**0.835**	0.257	-0.100	0.294	.763	.773	.859
PA AVE	**0.751**	-0.314	-0.515	-0.177	.662	.928	.959
PA STDEV	0.598	-0.360	-0.570	-0.328	**.488**	.813	.920
PA STDEVA	**0.802**	-0.208	-0.235	-0.046	.686	.741	.743

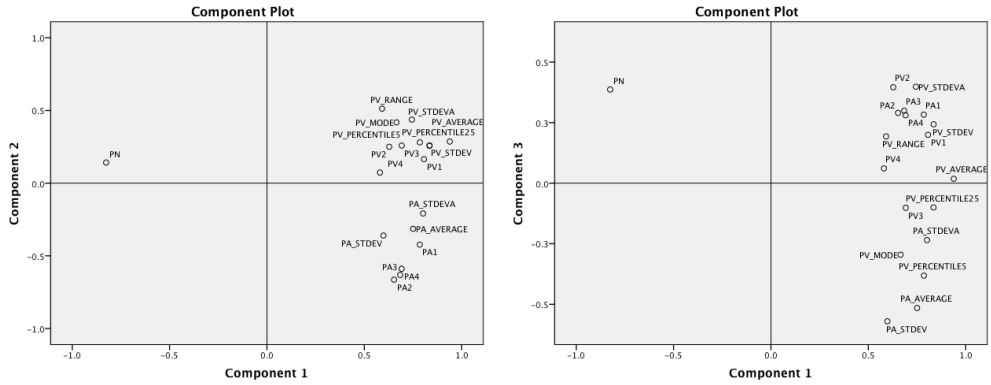


Figure 7.2.1 Loading plot of the principal components of pitch indices

The above results that Component 1 represents almost all the indices, while Component 2 mainly represents PA2 and PA3 and Component 3 mainly represents PA AVE and PA STDEV, may suggest one main dimension of the variance based on the current dataset. In other words, the pitch indices may form a single dimension for the samples used in this study.

7.2.3 Characteristics of the sound categories in terms of pitch features

Characteristics of the four sound categories, i.e. water, wind, birdsong, and urban, and differences among them are analysed in terms of the pitch indices with one-way analysis of variance (ANOVA) and with the principal components extracted in the last section as well as a number of key indices.

7.2.3.1 Comparison among the categories by the means of pitch indices with one-way analysis of variance

Table 7.2.6 Descriptives of pitch indices for the four categories

		N	Mean	Std. Deviation	Minimum	Maximum		N	Mean	Std. Deviation	Minimum	Maximum
Water	PV1	24	232	228	94	1078	PV AVE	34	233	85	143	513
Wind		9	530	193	94	707		23	202	63	113	366
Bird		28	2593	1288	473	4237		28	1754	503	475	2365
Urban		14	289	316	88	1196		17	221	102	123	507
Water	PV2	19	276	339	94	1592	PV Mode	34	103	47	90	350
Wind		3	261	289	94	595		23	102	54	90	350
Bird		28	1030	521	86	2174		28	1313	1008	90	2860
Urban		12	249	216	93	874		17	128	70	90	370
Water	PV3	14	199	146	93	557	PV STDEV	34	204	104	67	447
Wind		0	.	.	.	.		23	152	65	30	243
Bird		27	1055	671	162	2301		28	872	190	305	1272
Urban		9	178	136	78	472		17	176	134	39	462
Water	PV4	12	294	320	88	1262	PV STDEVA	34	206	103	68	447
Wind		0	.	.	.	.		23	155	62	50	260
Bird		26	860	605	123	3081		28	880	347	173	1310
Urban		9	304	364	93	949		17	153	111	40	385
Water	PA1	24	0.238	0.134	0.139	0.672	PV Range	34	2085	864	699	3651
Wind		9	0.242	0.129	0.128	0.496		23	1163	676	476	2912
Bird		28	1.434	0.533	0.484	2.605		28	3534	410	1911	3965
Urban		14	0.800	0.591	0.170	1.841		17	1513	1237	237	3595
Water	PA2	19	0.172	0.083	0.081	0.433	PV	34	83	3	80	93
Wind		3	0.132	0.018	0.114	0.149	Percentile5	23	83	3	80	94
Bird		28	0.872	0.419	0.360	2.025		28	466	255	88	891
Urban		12	0.599	0.505	0.142	1.735		17	87	5	81	97
Water	PA3	14	0.142	0.052	0.069	0.279	PV	34	105	20	93	178
Wind		0	.	.	.	.	Percentile25	23	101	28	93	225
Bird		27	0.698	0.321	0.187	1.776		28	1092	598	98	2109
Urban		9	0.528	0.507	0.131	1.682		17	114	23	94	175
Water	PA4	12	0.109	0.028	0.067	0.157	PA AVE	34	0.270	0.040	0.220	0.439
Wind		0	.	.	.	.		23	0.279	0.042	0.204	0.387
Bird		26	0.616	0.263	0.252	1.470		28	1.050	0.471	0.543	2.828
Urban		9	0.443	0.476	0.104	1.607		17	0.513	0.274	0.180	1.138
Water	PN	34	0.969	0.080	0.594	1.000	PA STDEV	34	0.081	0.025	0.055	0.167
Wind		23	0.816	0.286	0.151	1.000		23	0.102	0.038	0.063	0.199
Bird		28	0.352	0.248	0.027	0.866		28	0.557	0.439	0.230	2.197
Urban		17	0.649	0.373	0.040	1.000		17	0.241	0.191	0.044	0.831
Water							PA STDEVA	34	0.088	0.039	0.055	0.251
Wind								23	0.110	0.046	0.063	0.223
Bird								28	0.447	0.090	0.297	0.668
Urban								17	0.240	0.202	0.044	0.828

Table 7.2.7 Test of homogeneity of variances and ANOVA of pitch indices for the four categories

	Test of Homogeneity of Variances		ANOVA			Test of Homogeneity of Variances		ANOVA	
	Levene Statistic	Sig.	F	Sig.		Levene Statistic	Sig.	F	Sig.
PV1	31.52	0.000	45.92	0.000	PV AVE	28.98	0.000	213.04	0.000
PV2	4.71	0.005	17.26	0.000	PV Mode	201.98	0.000	34.85	0.000
PV3	14.62	0.000	17.98	0.000	PV STDEV	4.76	0.004	185.75	0.000
PV4	0.98	0.385	7.12	0.002	PV STDEVA	37.79	0.000	84.74	0.000
PA1	15.25	0.000	40.12	0.000	PV Range	13.85	0.000	42.53	0.000
PA2	8.02	0.000	15.50	0.000	PV Percentile5	46.54	0.000	55.25	0.000
PA3	8.09	0.001	14.07	0.000	PV Percentile25	103.31	0.000	66.67	0.000
PA4	5.71	0.006	13.04	0.000	PA AVE	20.34	0.000	50.84	0.000
PN	16.90	0.000	34.40	0.000	PA STDEV	15.73	0.000	23.03	0.000
					PA STDEVA	13.77	0.000	77.96	0.000

The mean values of the four categories are compared with one-way analysis of variance (ANOVA) in the software of SPSS Statistics, in terms of the remaining 19 pitch indices as discussed in Section 7.2.1, in order to examine if the sound categories differ from each other significantly in one or more indices. The descriptives of the indices for the four categories are shown in Table 7.2.6, including mean, standard deviation, minimum, and maximum.

Before the ANOVA, the assumption of ANOVA, i.e., the homogeneity of variances is firstly tested. Table 7.2.7 shows the results of the test for the indices. It can be seen that p value (Sig.) of all the indices are less than an alpha (α) level of 0.05, which means the assumption that the variances of the categories are equal is rejected; in other words, the variances are unequal. Table 7.2.7 also shows the results of ANOVA, in terms of F ratio and the significance of the F ratio (Sig.), i.e. p value, the specific meanings of which can be found Chapter 4 Section 4.1.1. It can be seen, for all the indices, that the significance of F ratio is less than an α level of 0.05, which reject the null hypothesis that all the means are equal. In other words, it suggests some significant differences in terms of all the indices among the means of the categories, or between at least two categories. However, since the variances are unequal, the results may be inaccurate, and thus further post hoc tests are checked for verifying and examining which of the specific categories differ.

Post hoc tests are based on the Dunnett's T3 method, as variances are unequal. The results of the post hoc tests are shown in Table 7.2.8, displaying the multiple comparisons among the categories, in terms of difference between the means of each category with each of the other three, where asterisks (*) indicate significantly different group means at an alpha level of 0.05. Here, post hoc tests are not performed for PV3, PV4, PA3 and PA4 because the wind sound group has no cases that show result in terms of the indices,

as shown in Table 7.2.6. Table 7.2.8 shows that these pitch indices, except for PV2 and PA2, all exhibit significant mean differences between birdsong and the other three categories of sounds, while PV2 shows significant differences between categories of birdsong and water and between birdsong and urban, PA2 shows significant differences between categories of birdsong and water and between birdsong and wind. In addition, PV1 and PV Range also show significant mean differences between categories of water and wind. PA1, PN, PV Percentile5, PA AVE, PA STDEV and PA STDEVA show significant mean differences between categories of water and urban. PA1, PA2, PA AVE and PA STDEV show significant mean differences between categories of wind and urban. Birdsongs have higher or much higher pitch values and pitch strengths than the other three categories in terms of means, and lower percentage of audible pitches over time (PN) than the other three (seen from both Table 7.2.6 and Table 7.2.8). Urban sounds generally have higher means of pitch strengths than water and wind sounds, and lower PN than water sounds. Water sounds have lower mean of PV1 and higher mean of PV Range than wind sounds.

In other words, the results show that there are certain significant differences among the categories in terms of the pitch indices. In general, for pitch values, birdsongs have higher mean values than the other three categories of sound. For pitch strengths, birdsongs have higher mean values than urban sounds, and than water and wind sounds. For percentage of audible pitches over time (PN), water and wind sounds have higher mean values than urban sounds, and than birdsongs. Characteristics of and differences among the categories are further discussed in the following section.

Table 7.2.8 Multiple comparisons of pitch indices for the four categories

Mean Difference (I-J)	(I) Category	Water			Wind			Bird			Urban		
Dependent Variable	(J) Category	Wind	Bird	Urban	Water	Bird	Urban	Water	Wind	Urban	Water	Wind	Bird
PV1	Dunnett T3	-298*	-2361*	-56	298*	-2063*	242	2361*	2063*	2305*	56	-242	-2305*
PV2	Dunnett T3	15	-754*	27	-15	-769	13	754*	769	781*	-27	-13	-781*
PA1	Dunnett T3	-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	Dunnett T3	0.040	-0.700*	-0.427	-0.040	-0.739*	-0.467*	0.700*	0.739*	0.272	0.427	0.467*	-0.272
PN	Dunnett T3	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	Dunnett T3	31	-1521*	12	-31	-1552*	-19	1521*	1552*	1533*	-12	19	-1533*
PV Mode	Dunnett T3	1	-1210*	-24	-1	-1211*	-25	1210*	1211*	1185*	24	25	-1185*
PV STDEV	Dunnett T3	52	-667*	29	-52	-720*	-24	667*	720*	696*	-29	24	-696*
PV STDEVA	Dunnett T3	51	-674*	53	-51	-724*	2	674*	724*	726*	-53	-2	-726*
PV Range	Dunnett T3	922*	-1449*	572	-922*	-2371*	-350	1449*	2371*	2021*	-572	350	-2021*
PV Percentile5	Dunnett T3	0	-383*	-4*	0	-382*	-4	383*	382*	379*	4*	4	-379*
PV Percentile25	Dunnett T3	4	-987*	-9	-4	-991*	-13	987*	991*	978*	9	13	-978*
PA AVE	Dunnett T3	-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	Dunnett T3	-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	Dunnett T3	-0.021	-0.359*	-0.151*	0.021	-0.337*	-0.130	0.359*	0.337*	0.207*	0.151*	0.130	-0.207*

7.2.3.2 Characteristics of the sound categories in terms of principal components and key pitch indices

Characteristics of the sound categories are analysed further visually with a number of key indices and the principal components extracted in Section 7.2.2. As indicated in Sections 7.2.1 and 7.2.2, high correlations exist among the pitch strengths of the best four average pitches for the whole duration, among the indices of pitch values over time, and among the indices of pitch strengths over time. Moreover, the indices can be grouped into three, i.e. the index of percentage of audible pitches over time (PN), the PV indices and the PA indices, based on the PCA. Thus, from the pitch values of the best four average pitches for the whole duration, pitch strengths of the best four average pitches, the percentage of audible pitches over time, statistic indices of pitch values over time and pitch strengths over time, one index is selected respectively as key index here. They are PV1, PA1, PN, PV AVE and PA AVE, which generally contribute most to the first component by PCA as shown in Table 7.2.5.

The 102 sound samples are thus plotted in a two-dimensional coordinate system with its axes presenting PV1 and PA1 and in another coordinate system presenting PV AVE and PA AVE. The plots are shown in Figure 7.2.2 (a) and (b) respectively, although it should be noted there are some correlations between the axes due to the correlations between the indices. In both plots, the recordings in birdsong category have relatively high pitch values and pitch strengths, 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, especially in Figure 7.2.2 (b), both with 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. Recordings in 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 between about 0 to 1 for PA AVE. These results suggest that there are certain differences in characteristics among the sound categories. Generally, water and wind sounds both have low pitch values and low pitch strengths; birdsongs have high pitch values and high pitch strengths; and urban sounds have low pitch values and a relatively wide range of pitch strengths.

In terms of percentage of audible pitches over time (PN), results of the recordings are shown in Figure 7.2.3, in which the minimum, first quartile (25th percentile), median (50th percentile), third quartile (75th percentile) and maximum scores of each of the 21 subcategories are presented. It shows that the PN values of all birdsong subcategories and of traffic subcategory in urban sounds are relatively low; those of all water sound subcategories and of church bells and fountain subcategories in urban sounds are high,

close to 1; while those of the rest subcategories, i.e. wind sounds and a majority of urban sounds, are in a median range. In other words, birdsongs and traffic sounds have fewer pitches audible over time than wind and some urban sounds, and than water, church bells and fountain sounds, which have audible pitches nearly all time over the duration (PN values of close to 1). Among the categories, generally, birdsongs have low values of PN and water sounds have high values. It is noted that the index of PN reflects a relative value here, i.e., relative audible pitches; although birdsongs have fewer audible pitches, the pitch amplitudes are higher than others.

In addition, based on the three principal components extracted in Section 7.2.2, a three-dimensional coordinate system is established with its axes presenting the components. As discussed in Section 7.2.2, Component 1 represents almost all the indices, including pitch values, pitch strengths and percentage of audible pitches; and Component 2 mainly represents pitch strengths of average pitches for the whole duration and Component 3 mainly represents average and standard deviation of pitch strengths over time. The 102 sound samples are correspondingly plotted in the coordinate system, as shown in Figure 7.2.4. In the figure, only the categories of water, birdsong and urban sound are displayed, since recordings in wind sound category do not show any results for a number of the indices and thus for the principal components. In Figure 7.2.4 (a) (Components 1 and 2), generally the recordings of water sound are located in the area where scores of Components 1 are below 0 and scores of Components 2 are above 0. The recordings of birdsong are located in the area where scores of Components 1 are above 0. The recordings of urban sound are located in the area where scores of Components 1 are below 0 and scores of Components 2 are both above are below 0 (mainly below 0). In Figure 7.2.4 (b) (Components 2 and 3), the recordings of birdsong scatter on the plane, while water and urban sound recordings, especially water sounds, gather in a line, which reflects the dependence of the scores of Components 2 and 3. These results, from another aspect, suggest again that water sounds have relatively low pitch values, low pitch strengths and high percentage of audible pitches; oppositely, birdsongs have relatively high pitch values, high pitch strengths and low percentage of audible pitches. Also, for water and urban sounds, pitch strengths of average pitches for the whole duration and pitch strengths over time are somehow correlative.

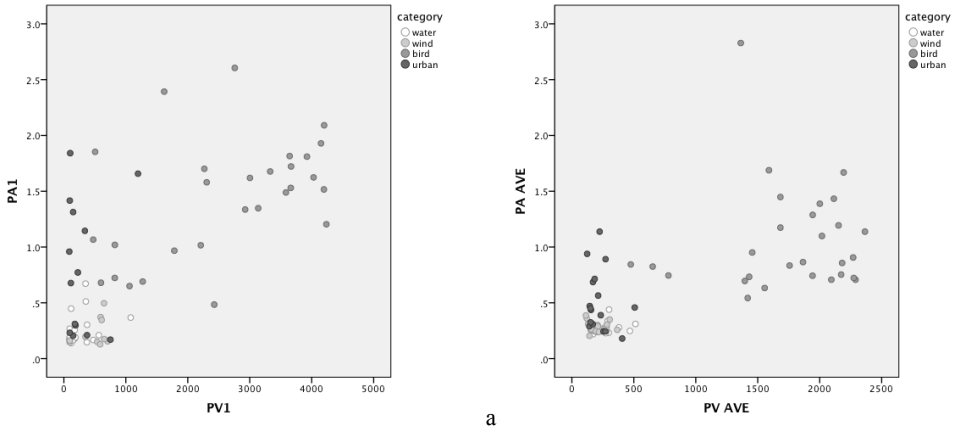


Figure 7.2.2 Characteristics of the four types of sound in terms of the key pitch indices

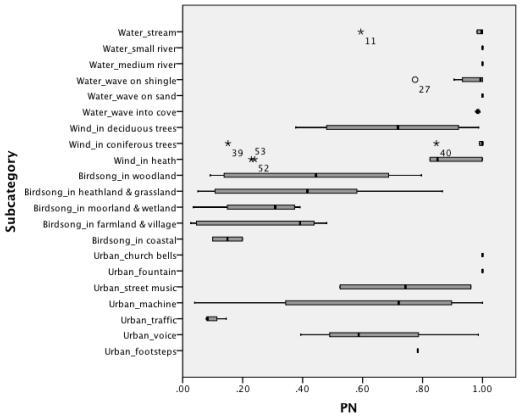


Figure 7.2.3 Characteristics of the sound subcategories in terms of percentage of audible pitches over time

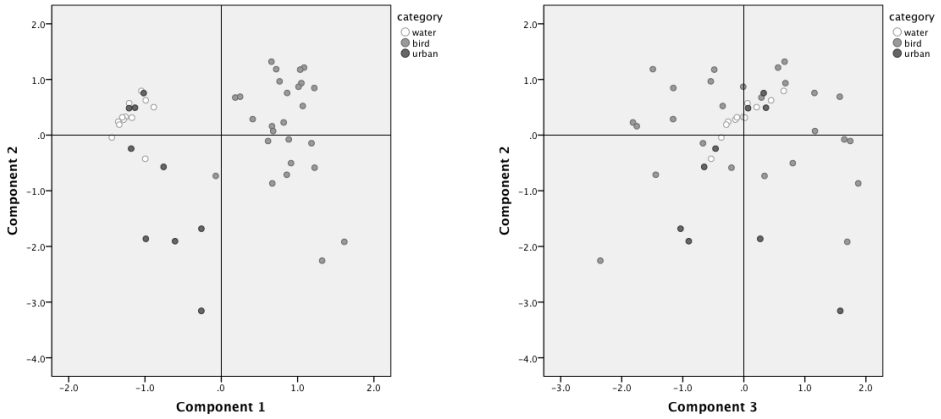


Figure 7.2.4 Characteristics of the four types of sound in terms of the principal components of pitch indices

In sum, based on both the analyses of ANOVA and plot displays with key indices and principal components, the results show certain differences among the four categories in terms of the pitch indices. In general, water sounds have low pitch values, low pitch strengths and high percentage of audible pitches; wind sounds have low pitch values and low pitch strengths; birdsongs have high pitch values, high pitch strengths and low percentage of audible pitches; and urban sounds have low pitch values and a relatively wide range of pitch strengths.

7.3 Characteristics of Natural and Urban Sounds in Soundscapes in Terms of Rhythm Features

In addition to the pitch features, characteristics of natural and urban sounds in soundscapes are analysed in terms of rhythm features. Firstly, similar to the analyses with pitch indices in Section 7.2, the correlations of the rhythm indices are examined to see whether some indices which are highly correlated, if so, some of them can be removed from the index set. Then, principal components of the rhythm indices are explored. Finally, the characteristics of natural and urban sounds are analysed, based on the principal components analysed.

7.3.1 Correlations between the rhythm indices

The correlations between the rhythm indices (see in Section 7.1) are examined based on the 102 samples, shown in Table 7.3.1, where the correlation coefficients which are higher than 0.8 are highlighted with bold numbers. The cases with missing value are excluded pairwise. It shows that generally the correlations are high (above 0.8) among the statistic indices of event interval by envelope method of event detection (IOI(E)), among those of attack slope (AS), among those of event interval by spectral flux method (IOI(SF)), and among those of spectral flux (SF). While between these different sets of indices, the correlations are not high, generally all lower than 0.6. In addition, the indices event density by envelope method (ED(E)) have medium correlations with indices of IOI(E) and AS, about 0.7 to 0.8. Indices of ED(SF) have medium correlations with indices of IOI(SF). The correlations between periodicity (BS) and all the other indices are not high, generally lower than 0.5.

Table 7.3.1 Pearson's correlations of rhythm indices

	IOI(E) AVE	IOI(E) Median	IOI(E) Mode	IOI(E) STDEV	IOI(E) Range	IOI(E) Percentile10	IOI(E) Percentile90	ED(E)	AS AVE	AS STDEV	AS Percentile10	AS Percentile90	IOI(SF) AVE	IOI(SF) Median	IOI(SF) Mode	IOI(SF) STDEV	IOI(SF) Range	IOI(SF) Percentile10	IOI(SF) Percentile90	ED(SF)	SF AVE	SF STDEV	SF Percentile10	SF Percentile90	BS
IOI(E) AVE	1																								
IOI(E) Median	.961**	1																							
IOI(E) Mode	.781**	.849**	1																						
IOI(E) STDEV	.916**	.804**	.581**	1																					
IOI(E) Range	.840**	.716**	.457**	.968**	1																				
IOI(E) Percentile10	.805**	.730**	.608**	.670**	.650**	1																			
IOI(E) Percentile90	.938**	.796**	.592**	.987**	.951**	.746**	1																		
ED(E)	-.676**	-.669**	-.567**	-.656**	-.695**	-.703**	-.759**	1																	
AS AVE	-.535**	-.532**	-.435**	-.493**	-.512**	-.617**	-.561**	.803**	1																
AS STDEV	-.226*	-.260**	-.207*	-0.161	-0.150	-.306**	-0.172	.277**	.598**	1															
AS Percentile10	-.619**	-.545**	-.425**	-.581**	-.575**	-.556**	-.586**	.867**	.878**	.332**	1														
AS Percentile90	-.482**	-.482**	-.335**	-.343**	-.321**	-.568**	-.383**	.648**	.885**	.462**	.703**	1													
IOI(SF) AVE	0.060	0.079	0.062	0.041	0.034	0.048	0.029	-0.111	0.186	.402**	0.026	0.126	1												
IOI(SF) Median	0.033	0.053	0.095	0.012	0.008	0.086	0.038	-0.108	0.151	.394**	-0.004	0.052	.904**	1											
IOI(SF) Mode	0.074	0.165	.316**	-0.003	-0.038	0.137	0.016	-0.067	0.162	0.194	0.038	0.154	.617**	.930**	1										
IOI(SF) STDEV	0.098	0.096	-0.060	0.072	0.052	-0.098	-0.026	-0.062	.230*	.410**	0.057	.224*	.739**	.464**	0.062	1									
IOI(SF) Range	0.087	0.089	-0.088	0.037	0.001	-0.173	-0.101	0.024	.255**	.251*	0.137	.301**	.442**	0.178	0.007	.865**	1								
IOI(SF) Percentile10	0.047	0.105	.226*	-0.008	-0.010	.203*	0.036	-0.102	0.065	0.065	-0.014	0.046	.758**	.948**	.897**	0.194	0.100	1							
IOI(SF) Percentile90	0.086	0.085	0.042	0.108	0.113	0.023	0.040	-0.101	0.130	.293**	0.026	0.166	.858**	.458**	0.187	.765**	.498**	.404**	1						
ED(SF)	-0.039	-0.031	0.061	-0.054	-0.051	0.098	0.005	0.015	-.327**	-.322**	-0.166	-.347**	-.607**	-.387**	-.251*	-.710**	-.683**	-.332**	-.656**	1					
SF AVE	0.119	0.128	0.066	0.085	0.052	-0.020	-0.060	-0.049	0.178	.233*	0.053	0.193	.380**	.321**	.271**	.380**	.305**	0.192	.228*	-.259**	1				
SF STDEV	-0.059	-0.070	-0.099	-0.044	-0.055	-0.151	-0.148	0.083	.335**	.430**	0.171	.273**	.604**	.482**	0.154	.592**	.406**	0.155	.436**	-.446**	.859**	1			
SF Percentile10	0.153	0.152	0.091	0.127	0.105	0.041	-0.007	-0.111	0.011	0.068	-0.054	0.060	0.160	0.125	0.160	0.166	0.156	0.108	0.112	-0.071	.966**	.788**	1		
SF Percentile90	0.047	0.047	0.007	0.034	0.011	-0.060	-0.111	0.035	.228*	.242*	0.124	.247*	.342**	0.179	0.192	.373**	.313**	0.169	.352**	-.294**	.957**	.975**	.893**	1	
BS	-.234*	-.233*	-0.163	-.211*	-.215*	-.251*	-.225*	.354**	.517**	0.127	.403**	.486**	-0.008	-0.023	0.119	0.070	0.171	0.050	0.040	-0.149	.248*	0.173	.206*	.255*	1

Among the highly correlated indices, the ones that have particularly high correlations with some others, higher than 0.95, are removed from the index set. For the indices of IOI(E), the correlations are particularly high between IOI(E) AVE and IOI(E) Median and among IOI(E) STDEV, IOI(E) Range and IOI(E) Percentile90, in which IOI(E) Median, IOI(E) Range and IOI(E) Percentile90 are removed from the index set, since their variance can be represented by the remaining indices (IOI(E) AVE and IOI(E) STDEV). For the SF indices, the correlations are particularly high between SF AVE and SF Percentile10, between SF AVE and SF Percentile90, and between SF STDEV and SF Percentile90, in which SF Percentile10 and SF Percentile90 are removed. For both the indices of AS and IOI(SF), there is no particularly high correlation between. These removed indices, i.e. IOI(E) Median, IOI(E) Range, IOI(E) Percentile90, SF Percentile10 and SF Percentile90, are greyed in Table 7.3.1 for illustrating.

7.3.2 Principal components of the rhythm indices

The principal components of the rhythm indices are analysed using PCA based on the results of the 102 recordings, in order to reduce the dimensionality of the dataset, since there are certain correlations between the indices as discussed in Section 7.3.1. As the rhythm indices are generally derived from two event detection methods, the analysis is firstly implemented based on all the remaining rhythm indices in Section 7.3.1, and then based on the indices from envelope method, and based on the indices from spectral flux method. The cases with missing value are excluded listwise.

7.3.2.1 Principal components of all the remaining rhythm indices

With the remaining 20 indices, the PCA is conducted on the correlation matrix of the indices. Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is taken before the PCA, the result of which is shown in Table 7.3.2. The result of 0.70 generally indicates the adequacy of the sample size and the availability of the analysis. From the 20 indices, 20 components are extracted, among which the eigenvalues of first five principal components are greater than one, as shown in Table 7.3.3. Seen from the table, which shows the variances that explained by the components, the first five components together explain 83.6% of the total variance, while the first three components together explain 71.2% of the total variance.

Table 7.3.2 KMO tests for PCAs based on rhythm indices

	Rhythm indices by both methods	Rhythm indices by E method	Rhythm indices by SF method
KMO Measure of Sampling Adequacy	0.701	0.726	0.682

Table 7.3.4 shows the correlations between the first five components and the indices, from which it can be seen that Component 1 has high correlations (above 0.6) with most indices based on envelope event detection method, which are IOI(E) AVE, IOI(E) Percentile10, ED(E), AS AVE, AS STDEV, AS Percentile10 and AS Percentile90; Component 2 has high correlations with a number of event interval indices, which are IOI(E) AVE, IOI(E) Mode, IOI(SF) AVE, IOI(SF) Median, IOI(SF) Mode and IOI(SF) Percentile10; Component 3 has high correlations with a number of event interval indices based on spectral flux event detection method, which are IOI(SF) Mode, IOI(SF) STDEV and IOI(SF) Percentile10; Component 4 has high correlations with spectral flux indices, which are SF AVE and SF STDEV; and Component 5 do not have any high correlation with any of the indices. Thus, Component 1 mainly represents event interval, event density and attack slope based on envelope event detection method; Component 2 mainly represents event interval (based on both the methods); Component 3 mainly represents event interval based on the spectral flux method; and Component 4 mainly represents spectral flux. These results can also be seen on the component loading plots, shown in Figure 7.3.1, where the first three components are displayed. The indices in the plots are generally clustered in groups, e.g. the IOI(E) indices, AS indices, IOI(SF) indices, and SF indices.

Table 7.3.4 also shows the proportion of each index's variance that can be explained by the retained principal components. When first 5 components are retained from the 20 components, the proportions of the indices' variance that can be explained are high except for BS, which has a value below 0.5, indicating that these indices are well represented by the principal components. When 4 components are retained, the indices have relatively high proportion values except for AS STDEV and BS. When 3 components are retained, AS STDEV, SF AVE, SF STDEV and BS, which have relatively low proportion values of below 0.5, are not well represented. Overall, the first 4 components are generally necessary to represent the original indices, which account for 78.0% of the total variance.

Table 7.3.3 Total variance explained by the components based on all the rhythm indices by both methods

Component	Eigenvalues	% of Variance	Cumulative %
1	**6.644**	**33.222**	**33.222**
2	**5.205**	**26.024**	**59.246**
3	**2.397**	**11.986**	**71.232**
4	**1.353**	**6.765**	**77.998**
5	**1.121**	**5.605**	**83.603**
6	0.826	4.132	87.734
7	0.540	2.700	90.434
8	0.480	2.398	92.832
9	0.402	2.012	94.845
10	0.317	1.584	96.429
11	0.238	1.188	97.617
12	0.163	0.813	98.430
13	0.102	0.512	98.942
14	0.067	0.334	99.277
15	0.044	0.219	99.496
16	0.039	0.193	99.689
17	0.025	0.127	99.816
18	0.020	0.101	99.918
19	0.013	0.067	99.985
20	0.003	0.015	100.000

Table 7.3.4 Component matrix and communalities for all the rhythm indices by both methods

	Component Matrix					Communalities		
	Component 1	Component 2	Component 3	Component 4	Component 5	Extraction of 3 components	Extraction of 4 components	Extraction of 5 components
IOI(E) AVE	**-0.660**	**0.622**	-0.071	-0.118	0.356	0.828	0.841	0.968
IOI(E) Mode	-0.474	**0.608**	0.120	-0.064	0.277	0.608	0.612	0.689
IOI(E) STDEV	-0.585	0.509	-0.173	-0.131	0.481	0.632	0.649	0.880
IOI(E) Percentile10	**-0.651**	0.516	0.001	0.019	0.187	0.690	0.691	0.726
ED(E)	**0.727**	-0.551	0.188	-0.058	-0.012	0.867	0.870	0.870
AS AVE	**0.910**	-0.192	0.165	-0.123	0.251	0.892	0.908	0.971
AS STDEV	**0.692**	0.030	0.035	-0.101	0.329	**0.481**	**0.492**	0.600
AS Percentile10	**0.758**	-0.361	0.199	-0.134	0.206	0.744	0.762	0.804
AS Percentile90	**0.829**	-0.079	0.102	-0.184	0.360	0.704	0.738	0.867
IOI(SF) AVE	0.466	**0.835**	-0.025	-0.117	-0.156	0.914	0.928	0.952
IOI(SF) Median	0.224	**0.734**	0.597	-0.066	-0.120	0.946	0.950	0.965
IOI(SF) Mode	0.219	**0.636**	**0.665**	-0.120	-0.163	0.895	0.909	0.935
IOI(SF) STDEV	0.532	0.485	**-0.625**	-0.127	-0.128	0.909	0.926	0.942
IOI(SF) Range	0.515	0.349	-0.597	-0.152	-0.109	0.743	0.766	0.778
IOI(SF) Percentile10	0.229	**0.690**	**0.605**	-0.053	-0.201	0.894	0.897	0.937
IOI(SF) Percentile90	0.457	0.584	-0.410	-0.075	-0.102	0.718	0.724	0.734
ED(SF)	-0.535	-0.526	0.333	0.271	0.077	0.673	0.747	0.753
SF AVE	0.369	0.450	0.022	**0.783**	0.085	**0.339**	0.952	0.959
SF STDEV	0.551	0.389	-0.179	**0.668**	0.050	**0.487**	0.933	0.936
BS	0.521	0.017	0.101	0.155	0.384	**0.282**	**0.306**	**0.453**

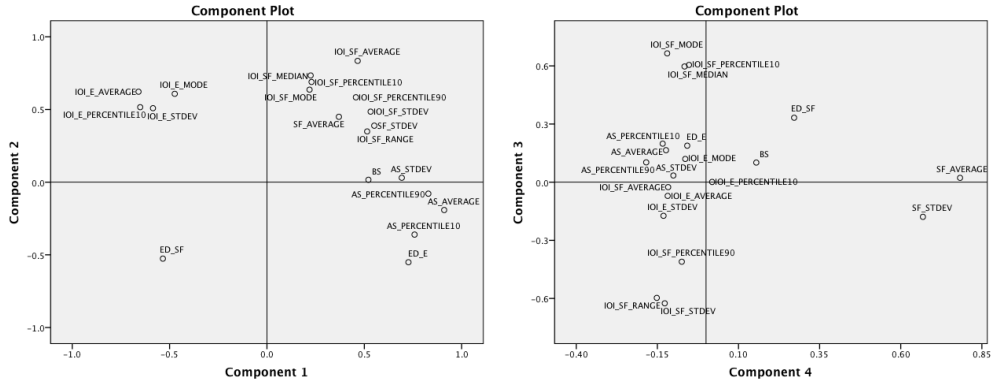


Figure 7.3.1 Loading plot of the principal components of rhythm indices by both methods

7.3.2.2 Principal components of the rhythm indices based on envelope method

Since as discussed in the last section that Component 1 represents almost all indices derived from envelope event detection method, principal components are explored more closely based respectively on the indices from envelope method and on the indices from spectral flux method in this and the next sections.

With the indices derived from envelope event detection method, the PCA is conducted on the correlation matrix of the 9 indices, from which 9 components are extracted. KMO measure of sampling adequacy shows a result of 0.73 (as seen in Table 7.3.2), which generally indicates the adequacy of the sample size and the availability of the analysis. The eigenvalues of first two principal components are greater than one, as shown in Table 7.3.5. The first component explains 63.2% and the second explains 15.5% of the total variance.

Table 7.3.6 shows the correlations between the first two principal components and the indices and the proportion of each index's variance that can be explained by the retained principal components. It can be seen that Component 1 has high correlations (above 0.6) with all the indices except for AS STDEV, and Component 2 has high correlation with AS STDEV. That is, Component 1 mainly represents event interval, event density and attack slope, and Component 2 mainly represents variation of attack slope. These results can also be seen on the component loading plots, shown in Figure 7.3.2 (a). The indices in the plots are clustered in two groups, which are the IOI(E) indices, and ED(E) and AS indices. When these two components are retained, generally all these indices are well represented.

Table 7.3.5 Total variance explained by the principal components based on the rhythm indices by envelope method and by spectral flux method

	Envelope			Spectral flux		
Component	Eigenvalues	% of Variance	Cumulative %	Eigenvalues	% of Variance	Cumulative %
1	**5.691**	**63.235**	**63.235**	**5.160**	**51.600**	**51.600**
2	**1.393**	**15.481**	**78.716**	**2.467**	**24.674**	**76.275**
3	0.745	8.281	86.998	**1.265**	**12.653**	**88.927**
4	0.435	4.837	91.835	0.538	5.375	94.302
5	0.380	4.225	96.060	0.323	3.233	97.536
6	0.204	2.271	98.331	0.098	0.985	98.520
7	0.097	1.077	99.409	0.078	0.783	99.304
8	0.031	0.342	99.751	0.042	0.416	99.719
9	0.022	0.249	100.000	0.024	0.242	99.961
10				0.004	0.039	100.000

Table 7.3.6 Component matrix and communalities for the rhythm indices by envelope method

	Component Matrix		Communalities
	Component 1	Component 2	Extraction of 2 components
IOI(E) AVE	**-0.891**	0.401	0.955
IOI(E) Mode	**-0.696**	0.396	0.642
IOI(E) STDEV	**-0.788**	0.453	0.826
IOI(E) Percentile10	**-0.824**	0.183	0.713
ED(E)	**0.918**	-0.060	0.846
AS AVE	**0.886**	0.423	0.964
AS STDEV	0.444	**0.606**	0.565
AS Percentile10	**0.849**	0.216	0.767
AS Percentile90	**0.753**	0.489	0.807

7.3.2.3 Principal components of the rhythm indices based on spectral flux method

For the indices derived from spectral flux event detection method, principal components of the 10 indices are explored. KMO measure of sampling adequacy shows a result of 0.68 (as seen in Table 7.3.2), which generally indicates the availability of the analysis. Among the 10 components extracted, the eigenvalues of first three principal components are greater than one, as shown in Table 7.3.5. They respectively explain 51.6%, 24.7% and 12.7% of the total variance.

Table 7.3.7 shows the correlations between the first two principal components and the indices and the proportion of each index's variance that can be explained by the retained principal components. Component 1 mainly represents event interval, event density and spectral flux, including all the indices except for IOI(SF) Mode; Component 2 mainly represents event interval, including IOI(SF) Median, IOI(SF) Mode and IOI(SF) Percentile10; and Component 3 mainly represents spectral flux, including SF AVE and SF STDEV. These results can also be seen on the component loading plots, shown in Figure 7.3.2 (b-d). In Figure 7.3.2 (d), the indices in the plots are clustered in four groups,

which are event density (ED(SF)), spectral flux (SF AVE and SF STDEV), a number of the IOI(SF) indices (IOI(SF) Median, IOI(SF) Mode and IOI(SF) Percentile10), and the rest IOI(SF) indices. When the first three components are retained, generally all these indices are well represented.

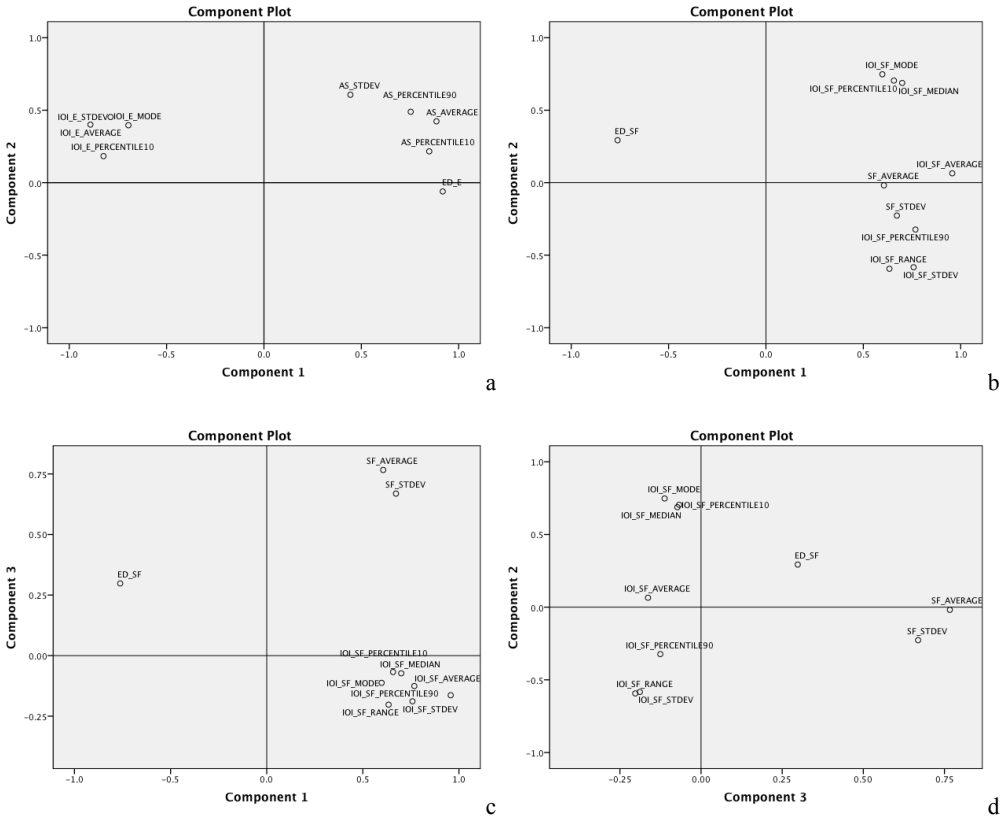


Figure 7.3.2 Loading plots of the principal components of the rhythm indices, (a) by envelope method, and (b), (c) and (d) by spectral flux method

Table 7.3.7 Component matrix and communalities for the rhythm indices by spectral flux method

	Component Matrix			Communalities
	Component 1	Component 2	Component 3	Extraction of 3 components extracted
IOI(SF) AVE	**0.957**	0.065	-0.164	0.947
IOI(SF) Median	**0.700**	**0.688**	-0.073	0.968
IOI(SF) Mode	0.598	**0.748**	-0.112	0.929
IOI(SF) STDEV	**0.758**	-0.583	-0.189	0.951
IOI(SF) Range	**0.634**	-0.593	-0.203	0.795
IOI(SF) Percentile10	**0.657**	**0.704**	-0.068	0.932
IOI(SF) Percentile90	**0.768**	-0.323	-0.126	0.710
ED(SF)	**-0.762**	0.293	0.298	0.755
SF AVE	**0.606**	-0.018	**0.767**	0.955
SF STDEV	**0.672**	-0.227	**0.669**	0.951

Based on the results of the PCAs with all rhythm indices, with indices derived from envelope event detection method, and with indices derived from spectral flux method, it can be summarised that there are three principal components or dimensions of the rhythm indices, more than those of pitch. Generally, the first principal component may represent event interval, event density and attack slope based on envelope method; the second may represent event interval and event density based on spectral flux method; and the third may represent spectral flux. Here, event interval and event density contribute to one component (for both envelope and spectral flux methods), as it may be expected that these two parameters be correlated conversely. Also, it can be seen from the results that though for each parameter a number of statistic indices are included, such as average, standard deviation, medium and percentiles, these statistic indices of a parameter mostly are in a single dimension, or say contribute to one component.

Based on the PCAs and correlations between the indices, a number of key indices can be identified, which generally contribute most to each of the principal components, i.e., have the highest correlations with each of the components as shown in Table 7.3.4, Table 7.3.6 and Table 7.3.7. They are IOI(E) AVE, ED(E), AS AVE, IOI(SF) AVE, ED(SF), SF AVE and BS, which represent the index groups of event interval, event density and attack slope based on envelope method, event interval, event density and spectral flux based on spectral flux method, and periodicity. The indices in each of the groups have high correlations among, and gather together in component loading spaces.

7.3.3 Characteristic of the sound categories in terms of rhythm features

Characteristics of the four sound categories, i.e. water, wind, birdsong, and urban, and differences among them are analysed in terms of the rhythm indices with one-way analysis of variance (ANOVA) and with the principal components extracted in the last section as well as a number of key indices.

7.3.3.1 Comparison among the categories by the means of rhythm indices with one-way analysis of variance

With the remaining 20 rhythm indices as discussed in Section 7.3.1, the mean values of the four categories are compared with one-way analysis of variance (ANOVA), in order to examine if the sound categories differ from each other significantly in one or more indices. The descriptives of the indices for the four categories are shown in Table 7.3.8, including mean, standard deviation, minimum, and maximum.

Table 7.3.8 Descriptives of rhythm indices for the four categories

		N	Mean	Std. Deviation	Minimum	Maximum		N	Mean	Std. Deviation	Minimum	Maximum
Water	IOI(E) AVE	34	1.164	0.611	0.302	2.402	IOI(SF) AVE	34	0.691	0.884	0.126	4.276
Wind		23	1.658	1.603	0.171	6.867		23	0.309	0.320	0.130	1.562
Bird		28	0.330	0.091	0.199	0.517		28	0.726	0.825	0.180	4.267
Urban		17	0.720	0.552	0.221	2.388		17	0.679	0.691	0.091	2.867
Water	IOI(E) Mode	33	0.8	0.7	0.2	3.0	IOI(SF)	34	0.469	0.870	0.108	4.373
Wind		19	0.7	0.8	0.1	3.6	Median	23	0.183	0.108	0.102	0.628
Bird		28	0.2	0.1	0.1	0.5		28	0.405	0.736	0.106	3.913
Urban		17	0.6	0.8	0.2	3.2		17	0.444	0.718	0.073	3.184
Water	IOI(E) STDEV	34	0.763	0.477	0.154	2.382	IOI(SF)	30	0.17	0.12	0.06	0.47
Wind		23	1.160	1.147	0.089	4.419	Mode	23	0.12	0.05	0.05	0.28
Bird		28	0.223	0.081	0.119	0.380		26	0.20	0.33	0.07	1.73
Urban		17	0.492	0.428	0.106	1.707		17	0.35	0.74	0.07	3.19
Water	IOI(E)	34	0.336	0.159	0.132	0.716	IOI(SF)	34	0.750	0.925	0.055	4.469
Wind	Percentile10	20	0.346	0.249	0.074	1.175	STDEV	23	0.395	0.628	0.064	2.230
Bird		28	0.119	0.041	0.065	0.279		28	0.975	0.712	0.121	3.141
Urban		17	0.234	0.143	0.098	0.651		17	0.806	0.944	0.038	3.437
Water	ED(E)	34	1.20	0.80	0.43	3.27	IOI(SF)	34	2.849	2.458	0.258	10.014
Wind		23	1.43	1.50	0.17	5.83	Range	23	1.888	2.894	0.309	10.402
Bird		28	3.24	0.90	1.87	5.03		28	4.380	2.354	0.696	9.903
Urban		17	2.01	1.06	0.43	4.53		17	3.776	4.822	0.202	20.274
Water	AS AVE	34	1.01	0.71	0.28	3.38	IOI(SF)	33	0.105	0.105	0.055	0.672
Wind		23	0.85	0.85	0.20	3.39	Percentile10	23	0.073	0.014	0.050	0.109
Bird		28	3.17	0.94	1.64	5.63		26	0.084	0.023	0.061	0.174
Urban		17	1.83	0.95	0.31	3.47		17	0.138	0.144	0.056	0.664
Water	AS STDEV	34	1.57	3.03	0.16	12.84	IOI(SF)	33	1.468	1.666	0.205	6.201
Wind		23	0.60	0.31	0.24	1.33	Percentile90	23	0.777	1.326	0.224	6.484
Bird		28	2.36	1.51	0.93	7.68		26	1.562	1.403	0.280	5.579
Urban		17	1.82	1.98	0.24	8.33		17	1.438	1.754	0.129	7.483
Water	AS	34	0.26	0.14	0.05	0.64	ED(SF)	34	3.19	2.27	0.17	7.87
Wind	Percentile10	20	0.33	0.47	0.04	1.67		23	4.89	2.35	0.47	7.70
Bird		28	1.01	0.51	0.44	2.80		28	2.15	1.29	0.20	5.53
Urban		17	0.51	0.31	0.06	1.17		17	2.98	2.76	0.37	10.97
Water	AS	34	2.08	1.88	0.62	12.13	SF AVE	34	20.9	18.2	4.0	92.8
Wind	Percentile90	20	1.91	1.24	0.71	5.11		23	28.5	23.0	3.5	80.6
Bird		28	5.45	1.31	2.89	7.63		28	33.2	25.1	3.9	86.2
Urban		17	3.33	1.48	0.74	5.90		17	61.4	40.5	8.0	135.2
Water	BS	34	0.00	0.00	0.00	0.00	SF STDEV	34	6.3	10.6	1.0	61.4
Wind		23	0.00	0.00	0.00	0.00		23	6.8	6.9	0.5	24.9
Bird		28	0.54	0.51	0.00	1.00		28	14.2	12.8	1.4	46.3
Urban		17	0.41	0.51	0.00	1.00		17	21.2	18.4	2.1	78.3

The assumption of ANOVA, i.e. the homogeneity of variances is firstly tested before the ANOVA, the results of which are displayed in Table 7.3.9. It shows that p value (Sig.) of ED(E), AS AVE, AS Percentile90, IOI(SF) AVE, IOI(SF) Median, IOI(SF) STDEV, IOI(SF) Range and IOI(SF) Percentile90 are greater than an alpha (α) level of 0.05, which means it fails to reject the null hypothesis, that is, the variances of the categories are equal and the assumption of homogeneity of variance has been met. For the other indices, the p values are less than an α level of 0.05, then the hypothesis that the variances are equal is rejected, that is, the variances are unequal.

Table 7.3.9 Test of homogeneity of variances and ANOVA of rhythm indices for the four categories

	Test of Homogeneity of Variances		ANOVA			Test of Homogeneity of Variances		ANOVA	
	Levene Statistic	Sig.	F	Sig.		Levene Statistic	Sig.	F	Sig.
IOI(E) AVE	15.673	0.000	10.900	0.000	IOI(SF) AVE	1.800	0.152	1.672	0.178
IOI(E) Mode	6.813	0.000	4.881	0.003	IOI(SF) Median	1.830	0.147	0.848	0.471
IOI(E) STDEV	17.199	0.000	9.864	0.000	IOI(SF) Mode	4.072	0.009	1.519	0.215
IOI(E) Percentile10	7.747	0.000	12.129	0.000	IOI(SF) STDEV	0.960	0.415	2.189	0.094
ED(E)	2.367	0.076	21.278	0.000	IOI(SF) Range	1.370	0.257	3.189	0.027
AS AVE	1.851	0.143	42.959	0.000	IOI(SF) Percentile10	3.841	0.012	2.178	0.096
AS STDEV	4.334	0.007	3.028	0.033	IOI(SF) Percentile90	1.279	0.286	1.296	0.280
AS Percentile10	6.262	0.001	23.204	0.000	ED(SF)	3.726	0.014	6.907	0.000
AS Percentile90	0.519	0.670	30.318	0.000	SF AVE	5.976	0.001	9.488	0.000
BS	1217.220	0.000	18.196	0.000	SF STDEV	3.906	0.011	7.173	0.000

Table 7.3.10 Multiple comparisons of rhythm indices for the four categories

Mean Difference (I-J)	(I) Category	Water			Wind			Bird			Urban		
Dependent Variable	(J) Category	Wind	Bird	Urban	Water	Bird	Urban	Water	Wind	Urban	Water	Wind	Bird
IOI(E) AVE	Dunnett T3	-0.494	0.834*	0.444	0.494	1.328*	0.938	-0.834*	-1.328*	-0.390	-0.444	-0.938	0.390
IOI(E) Mode	Dunnett T3	0.1	0.6*	0.2	-0.1	0.5	0.2	-0.6*	-0.5	-0.4	-0.2	-0.2	0.4
IOI(E) STDEV	Dunnett T3	-0.397	0.541*	0.271	0.397	0.937*	0.668	-0.541*	-0.937*	-0.269	-0.271	-0.668	0.269
IOI(E) Percentile10	Dunnett T3	-0.01	0.217*	0.10	0.01	0.227*	0.11	-0.217*	-0.227*	-0.115*	-0.10	-0.11	0.115*
ED(E)	Tukey HSD	-0.23	-2.04*	-0.81	0.23	-1.81*	-0.58	2.04*	1.81*	1.23*	0.81	0.58	-1.23*
AS AVE	Tukey HSD	0.16	-2.16*	-0.82*	-0.16	-2.32*	-0.98*	2.16*	2.32*	1.34*	0.82*	0.98*	-1.34*
AS STDEV	Dunnett T3	0.97	-0.79	-0.25	-0.97	-1.75*	-1.22	0.79	1.75*	0.53	0.25	1.22	-0.53
AS Percentile10	Dunnett T3	-0.07	-0.75*	-0.25*	0.07	-0.68*	-0.18	0.75*	0.68*	0.50*	0.25*	0.18	-0.50*
AS Percentile90	Tukey HSD	0.17	-3.37*	-1.25*	-0.17	-3.54*	-1.41*	3.37*	3.54*	2.12*	1.25*	1.41*	-2.12*
IOI(SF) AVE	Tukey HSD	0.382	-0.035	0.012	-0.382	-0.417	-0.370	0.035	0.417	0.047	-0.012	0.370	-0.047
IOI(SF) Median	Tukey HSD	0.286	0.063	0.024	-0.286	-0.222	-0.261	-0.063	0.222	-0.039	-0.024	0.261	0.039
IOI(SF) Mode	Dunnett T3	0.05	-0.03	-0.18	-0.05	-0.08	-0.24	0.03	0.08	-0.15	0.18	0.24	0.15
IOI(SF) STDEV	Tukey HSD	0.355	-0.224	-0.055	-0.355	-0.579	-0.410	0.224	0.579	0.169	0.055	0.410	-0.169
IOI(SF) Range	Tukey HSD	0.962	-1.530	-0.926	-0.962	-2.492*	-1.888	1.530	2.492*	0.604	0.926	1.888	-0.604
IOI(SF) Percentile10	Dunnett T3	0.033	0.021	-0.032	-0.033	-0.011	-0.065	-0.021	0.011	-0.054	0.032	0.065	0.054
IOI(SF) Percentile90	Tukey HSD	0.691	-0.094	0.030	-0.691	-0.784	-0.660	0.094	0.784	0.124	-0.030	0.660	-0.124
ED(SF)	Dunnett T3	-1.69	1.04	0.21	1.69	2.74*	1.91	-1.04	-2.74*	-0.83	-0.21	-1.91	0.83
SF AVE	Dunnett T3	-7.7	-12.3	-40.5*	7.7	-4.6	-32.8*	12.3	4.6	-28.2	40.5*	32.8*	28.2
SF STDEV	Dunnett T3	-0.5	-7.9	-14.9*	0.5	-7.4	-14.4*	7.9	7.4	-7.0	14.9*	14.4*	7.0
BS	Dunnett T3	0.00	-0.54*	-0.41*	0.00	-0.54*	-0.41*	0.54*	0.54*	0.12	0.41*	0.41*	-0.12

Table 7.3.9 also shows the results of ANOVA, in terms of F ratio and its p value. It can be seen that the significance of F ratio is greater than an α level of 0.05 for the IOI(SF) indices, including IOI(SF) AVE, IOI(SF) Median, IOI(SF) Mode, IOI(SF) STDEV, IOI(SF) Percentile10 and IOI(SF) Percentile90, which suggests that there may be no

statistically significant difference in the means of the indices among the four categories; for the other indices, the significance of F ratio is less than 0.05, which reject the null hypothesis that all the means are equal. In other words, it suggests some significant differences in terms of these indices among the means of the categories, or between at least two categories, whereas for most of the indices the assumption of homogeneity of variances is not met, post hoc tests are thus further checked for verifying and examining which of the specific categories differ.

The results of the post hoc tests are shown in Table 7.3.10, by either Tukey's HSD or Dunnett's T3 method according to the homogeneity of variances of the categories in the index, i.e., Tukey's HSD is used for the indices with equal variances between categories, while Dunnett's T3 is used for the indices with unequal variances. It shows the multiple comparisons among the categories, in terms of difference between the means of each category with each of the other three, where asterisks (*) indicate significantly different group means at an α level of 0.05. From Table 7.3.10, it can be seen, the same as the results of ANOVA above, that there are significant mean differences between at least two categories in terms of all the indices except for most of the IOI(SF) indices. Among the IOI(E) indices, IOI(E) AVE and IOI(E) STDEV show significant mean differences between categories of bird and water, and between bird and wind; IOI(E) Mode only shows significant difference between categories of bird and water; and IOI(E) Percentile10, as well as ED(E), show significant differences between categories of bird and water, between bird and wind, and also between bird and urban. Birdsongs have lower mean values of IOI(E) and thus higher mean value of ED(E) than water, wind and urban sounds. Among the AS indices, AS AVE, AS Percentile10 and AS Percentile90 show significant differences between bird and the other three categories, in which AS AVE and AS Percentile90 also show significant differences between urban and water and between urban and wind, AS Percentile10 also shows significant difference between urban and water. AS STDEV shows significant difference between categories of bird and wind. Birdsongs generally have higher mean values of AS than urban sounds, and than water and wind sounds. Among the IOI(SF) and ED(SF) indices, IOI(SF) Range and ED(SF) show significant differences between categories of bird and wind. Birdsongs have higher mean values of IOI(SF) Range and lower mean values of ED(SF) than wind sounds. In terms of SF indices, SF AVE and SF STDEV show significant differences between categories of urban and water and between urban and wind. Urban sounds have higher mean value of SF than water and wind sounds. BS shows significant differences between categories of bird and water, between bird and wind, between urban and water, and between urban and wind. Birdsongs and urban sounds have higher BS than water and wind sounds.

In other words, the results show, in general, that birdsongs have lower mean values of event interval (IOI(E)) and thus higher mean value of event density (ED(E)) based on envelope event detection method than the other three categories of sound, while birdsongs have lower mean value of event density (ED(SF)) based on spectral flux event detection method than wind sounds. Event interval indices (IOI(SF)) based on spectral flux event detection method generally do not show any significant mean differences between the categories. Birdsongs generally have higher mean values of attack slope (AS) than urban sounds, and than water and wind sounds, while urban sounds have higher mean values of spectral flux (SF) than water and wind sounds. A number of birdsongs and urban sounds show periodicity (BS), while water and wind sounds do not. Characteristics of and differences among the categories are further discussed in the following section.

7.3.3.2 Characteristics of the sound categories in terms of principal components and key rhythm indices

Characteristics of the sound categories are analysed visually with a number of key indices and the principal components extracted in Section 7.3.2. With the key indices, which are IOI(E) AVE, ED(E), AS AVE, IOI(SF) AVE, ED(SF), SF AVE and BS as discussed in in Section 7.3.2, the 102 sound samples are plotted in two three-dimensional coordinate systems, with their axes respectively presenting IOI(E) AVE, ED(E) and AS AVE based on envelope event detection method, shown in Figure 7.3.3 (a) and (b), and presenting IOI(SF) AVE, ED(SF) and SF AVE based on spectral flux event detection method, shown in Figure 7.3.3 (c) and (d). In Figure 7.3.3 (a) and (b), based on envelope event detection method, recordings in water and wind sound categories have relatively high values of IOI(E) AVE, low values of ED(E), and low values of AS AVE; recordings in birdsong category have relatively low values of IOI(E) AVE, high values of ED(E), and high values of AS AVE; and recordings in urban sound category have a relatively wide range of values of IOI(E) AVE, ED(E) and AS AVE. Based on spectral flux event detection method, in Figure 7.3.3 (c) and (d), recordings in the four categories generally mix.

With the first four principal components extracted in Section 7.3.2 of all the remaining rhythm indices, the 102 sound samples are plotted in a four-dimensional coordinate system, with its axes presenting the components, as shown in Figure 7.3.4. As discussed in Section 7.3.2, Component 1 mainly represents event interval, event density and attack slope based on envelope event detection method; Component 2 mainly represents event interval based on both the methods; Component 3 mainly represents event interval based on the spectral flux method; and Component 4 mainly represents spectral flux. In Figure 7.3.4 (a) (Components 1 and 2), generally the recordings of water

sound and wind sound are located in the area where scores of Component 1 are below 0, while the recordings of birdsong are located in the area where scores of Components 1 are above 0. The recordings of urban sound are located in the area where scores of Components 1 are both above are below 0. In Figure 7.3.4 (b) (Components 3 and 4), the located areas of the recordings of four categories overlap, among which urban sound recordings have slightly higher scores of Components 4. These results, from another aspect, suggest again that water and wind sounds have relatively low event density (high event interval) and low attack slope based on envelope event detection method; oppositely, birdsongs have relatively high event density (low event interval) and high attack slope. Urban sounds have relatively higher mean values of spectral flux. The indices based on spectral flux event detection method, which mainly contribute to Components 2 and 3, generally do not show any differences among the categories.

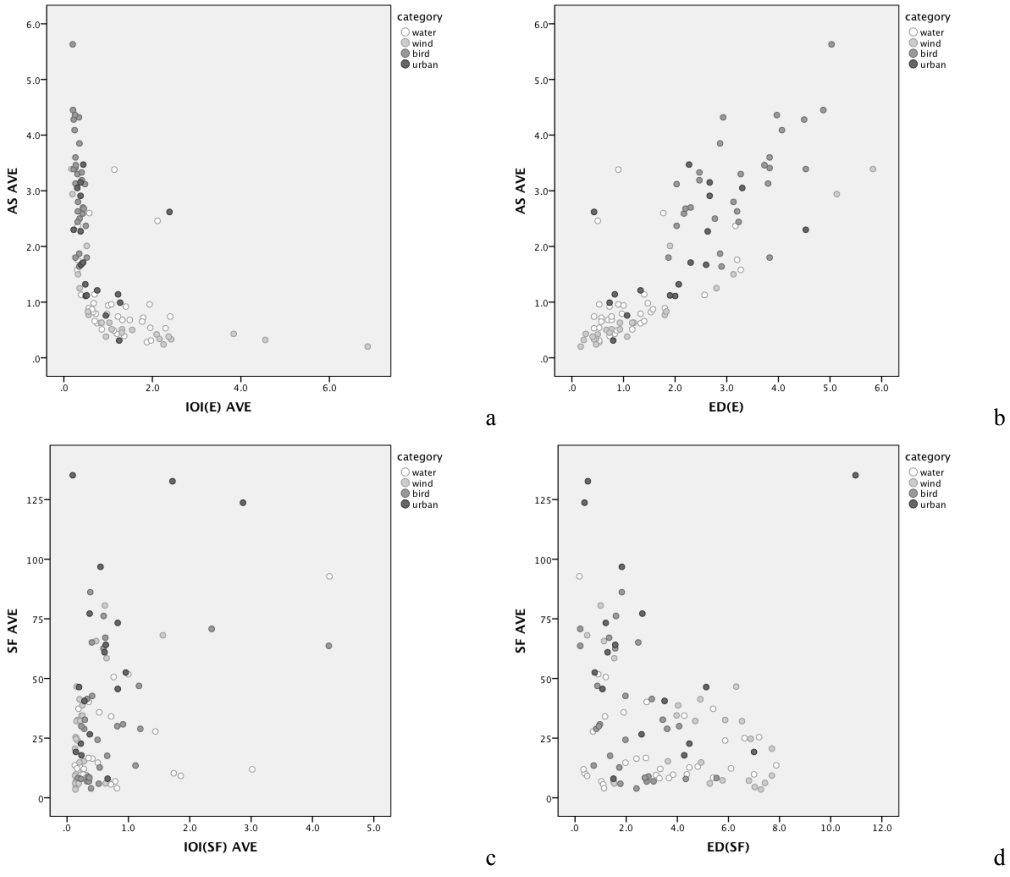
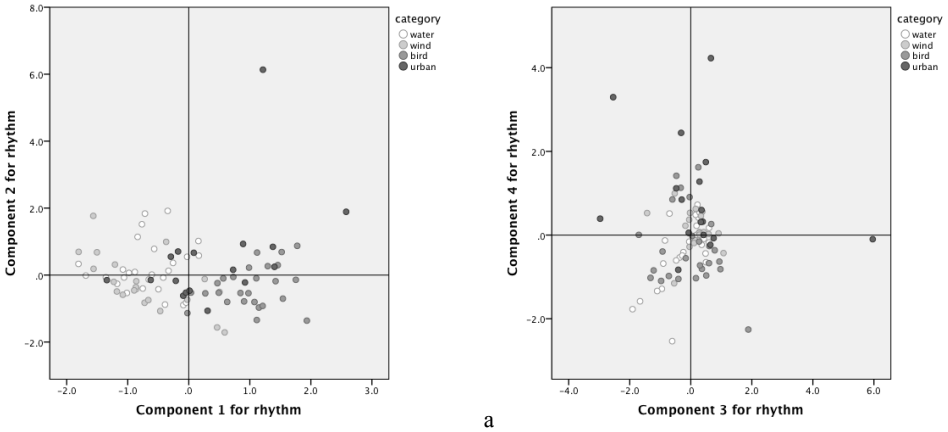


Figure 7.3.3 Characteristics of the four types of sound in terms of the key rhythm indices



a b

Figure 7.3.4 Characteristics of the four types of sound in terms of the principal components of rhythm indices

In sum, based on envelope event detection method, generally, water and wind sounds both have relatively high event interval, low event density, and low attack slope; oppositely, birdsongs have relatively low event interval, high event density, and high attack slope; and urban sounds have a relatively wide range of event interval, event density and attack slope. The indices based on spectral flux event detection method generally do not show any significant differences between the categories. A number of birdsongs and urban sounds show periodicity, while water and wind sounds do not.

7.4 Automatic Identification of Sound Categories with Pitch and Rhythm Indices

In this section, the correlations between and principal components of the pitch and rhythm indices are firstly examined. Then, discriminant function analyses are excuted to examine the ability of the pitch and rhythm indices to automatic identification of the four sound categories, which also reflect the characteristics of the sound categories.

7.4.1 Correlations between the pitch and the rhythm indices

The correlations between the pitch and rhythm indices are examined with the 102 sound recordings, shown in Table 7.4.1, where ** and * respectively indicate correlation is significant at the 0.01 level and 0.05 level (2-tailed). The correlation coefficients which are higher than 0.6 are highlight with bold numbers. It shows that the correlations

between the pitch indices and the rhythm indices are generally not high, the majority of which are not higher than 0.6. It suggests that the pitch indices set and the rhythm indices set do not share much common variance. The relatively high correlations (with coefficients between about 0.6 and 0.8) between the rhythm indices of event density based on envelope event detection method (ED(E)) and attack slope (AS) and the pitch indices may be interpreted as the correlations linked by the recordings in the current study, such as the birdsong recordings have high event density and attack slope based on envelope event detection method and meanwhile high pitch values and strengths and low percentage of audible pitches as discussed above in this chapter.

7.4.2 Principal components of the pitch and rhythm indices

Principal component analysis (PCA) is used in this section to reveal the dimensions of the pitch and rhythm indices, based on both key pitch and rhythm indices as discussed above, also shown in the left column of Table 7.4.3. Before implementation of the PCA, Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy shows a result of 0.79, which generally indicates the adequacy of the sample size and the availability of the analysis.

From the 12 indices or variables, 12 components are extracted. Table 7.4.2 shows the variances that explained by the components. Table 7.4.3 shows the correlations between the first three components and indices, and also shows the proportion of each index's variance that can be explained by the first three principal components. From the tables, it can be seen that the PCA show three principal components based on the key indices of both pitch and rhythm, as the eigenvalues of first three components are greater than one. The principal components have high correlations (above 0.6) with a number of the indices respectively. The results suggests Component 1 mainly represents pitch value, pitch strength, percentage of audible pitches over time (PV, PA, PN) and event interval, event density and attack slope based on envelope method (IOI(E) AVE, ED(E) and AS AVE). Component 2 mainly represents event interval and event density based on spectral flux method (IOI(SF) AVE and ED(SF)). Component 3 mainly represents spectral flux (SF AVE). These results can also be seen on the component loading plots, shown in Figure 7.4.1, where the first three components are displayed. When the first three components are retained, the proportions of all the indices' variance that can be explained are generally high (above 0.5), which means that generally all these indices are well represented by the principal components. The first three components together account for 73.4% of the total variance as shown in Table 7.4.2.

Table 7.4.1 Correlations between the pitch and rhythm indices

	PV1	PV2	PV3	PV4	PA1	PA2	PA3	PA4	PN	PV AVE	PV Mode	PV STDEV	PV STDEVA	PV Range	PV Percentile5	PV Percentile25	PA AVE	PA STDEV	PA STDEVA
IOI(E) AVE	-.406**	-.373**	-.355*	-0.222	-.437**	-.294*	-0.084	-0.085	.483**	-.385**	-.321**	-.349**	-.322**	-.286**	-.344**	-.352**	-.380**	-.335**	-.474**
IOI(E) Mode	-.347**	-.293*	-.311*	-0.077	-.298*	-0.083	0.076	0.126	.445**	-.328**	-.276**	-.284**	-.259*	-0.165	-.294**	-.306**	-.312**	-.290**	-.396**
IOI(E) STDEV	-.304**	-.311*	-0.239	-0.209	-.381**	-.274*	-0.131	-0.166	.431**	-.357**	-.302**	-.330**	-.305**	-.306**	-.322**	-.326**	-.364**	-.312**	-.456**
IOI(E) Percentile10	-.488**	-.429**	-.493**	-.300*	-.428**	-.311*	-0.151	-0.134	.583**	-.480**	-.382**	-.446**	-.407**	-.295**	-.415**	-.429**	-.419**	-.380**	-.510**
ED(E)	.533**	.401**	.505**	.359*	.497**	.377**	0.221	0.237	**-.676****	**.607****	.584**	.511**	.492**	.363**	**.618****	.579**	.583**	.502**	**.646****
AS AVE	**.612****	.464**	.479**	.324*	**.640****	.519**	.515**	.493**	**-.643****	**.736****	**.635****	**.661****	**.634****	.508**	**.694****	**.680****	**.728****	**.629****	**.801****
AS STDEV	.281*	.271*	0.176	0.069	.349**	0.225	0.192	0.123	-.235*	.240*	.199*	.232*	.208*	0.175	.220*	.197*	.241*	.203*	.253*
AS Percentile10	.508**	.323*	.369**	.288*	.502**	.421**	.367**	.342*	**-.634****	**.621****	**.613****	.545**	.497**	.395**	**.672****	.568**	**.740****	**.661****	**.740****
AS Percentile90	.569**	.387**	.414**	0.276	.589**	.456**	.463**	.437**	-.552**	**.696****	.553**	**.624****	**.610****	.471**	**.615****	**.651****	**.636****	.533**	**.720****
IOI(SF) AVE	0.112	0.124	-0.007	0.029	0.194	0.195	.313*	0.272	-0.063	0.095	0.035	0.123	0.102	0.138	0.024	0.077	0.044	-0.009	0.036
IOI(SF) Median	-0.028	-0.006	-0.118	-0.004	0.093	0.138	0.245	0.237	0.020	0.001	0.002	0.016	0.002	0.055	-0.014	-0.002	0.005	-0.034	-0.013
IOI(SF) Mode	-0.069	-0.041	-0.164	0.021	0.145	0.230	.320*	.351*	0.065	-0.015	-0.043	0.026	0.024	0.071	-0.047	-0.059	0.016	-0.020	0.035
IOI(SF) STDEV	.370**	.408**	.303*	0.136	.366**	.323*	.441**	.352*	-0.176	.211*	0.063	.243*	.237*	.201*	0.085	0.177	0.094	0.050	0.117
IOI(SF) Range	.408**	.462**	.411**	0.181	.414**	.398**	.495**	.413**	-.258**	.277**	0.143	.260**	.265**	.221*	0.189	.268**	.197*	0.150	.208*
IOI(SF) Percentile10	-0.171	-0.174	-0.213	0.007	0.000	0.089	.318*	.361*	0.137	-0.085	-0.055	-0.088	-0.068	0.038	-0.073	-0.074	-0.043	-0.065	-0.051
IOI(SF) Percentile90	0.195	0.126	0.099	0.119	.238*	0.141	.326*	0.276	-0.064	0.136	0.029	0.170	0.155	0.165	0.026	0.120	0.011	-0.021	0.006
ED(SF)	-.297**	-.296*	-0.147	-0.036	-.376**	-.323*	-.416**	-.348*	0.193	-.302**	-0.185	-.337**	-.324**	-.391**	-0.191	-.261**	-.221*	-0.152	-.260**
SF AVE	0.016	0.090	-0.099	-0.093	0.210	0.212	.347*	.346*	-0.056	0.026	0.034	0.057	0.065	-0.059	-0.034	0.009	0.060	0.029	0.147
SF STDEV	0.182	.273*	0.022	-0.066	.332**	.286*	.287*	0.244	-.210*	0.157	0.122	.209*	0.187	0.068	0.063	0.114	0.145	0.090	.239*
BS	.312**	0.178	.322*	0.010	.492**	.413**	.463**	.470**	-.336**	.461**	.411**	.367**	.416**	.256**	.504**	.511**	.493**	.402**	.578**

Table 7.4.2 Total variance explained by the components based on pitch and rhythm indices

Component	Eigenvalues	% of Variance	Cumulative %
1	**5.943**	**49.529**	**49.529**
2	**1.747**	**14.554**	**64.083**
3	**1.122**	**9.351**	**73.435**
4	0.813	6.774	80.209
5	0.623	5.188	85.397
6	0.528	4.398	89.796
7	0.373	3.109	92.905
8	0.342	2.851	95.756
9	0.200	1.663	97.419
10	0.143	1.194	98.613
11	0.105	0.874	99.488
12	0.061	0.512	100.000

Table 7.4.3 Component matrix and communalities for pitch and rhythm indices

	Component Matrix			Communalities
	Component 1	Component 2	Component 3	Extraction of 3 components
PV1	**0.794**	0.054	-0.214	0.679
PA1	**0.817**	0.193	0.079	0.712
PN	**-0.876**	0.012	0.113	0.780
PV AVE	**0.874**	-0.025	-0.190	0.801
PA AVE	**0.833**	-0.096	-0.018	0.704
IOI(E) AVE	**-0.675**	0.277	-0.034	0.533
ED(E)	**0.816**	-0.367	0.011	0.800
AS AVE	**0.903**	-0.036	0.093	0.826
IOI(SF) AVE	0.151	**0.862**	-0.110	0.777
ED(SF)	-0.367	**-0.747**	0.289	0.776
SF AVE	0.117	0.427	**0.793**	0.824
BS	0.558	-0.042	0.536	0.600

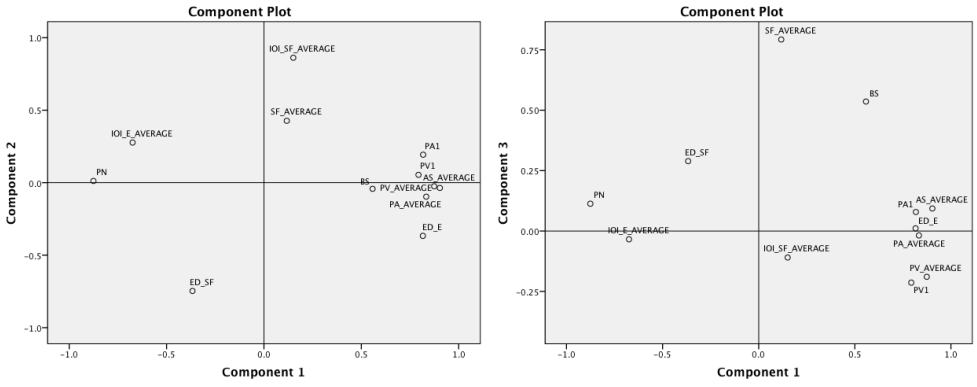


Figure 7.4.1 Loading plot of the principal components of pitch and rhythm indices

7.4.3 Discriminant function analysis based on the pitch and rhythm indices

Discriminant function analyses (DFAs) are used in this section to predict group membership through discriminant functions, i.e. linear combinations of the variables. With the 102 recordings with both known group membership and the values of variables, models are obtained to allow prediction of group membership with only the known variables. Also, discriminant functions give insight into the relationship between group membership and the variables used to predict group membership (Stockburger 2013).

7.4.3.1 Discriminant function analysis based on pitch indices

DFA is implemented firstly based on the remaining pitch indices. Here, for a number of the indices, the recordings or cases which do not show any results are assigned with a value of zero, e.g. the wind sound recordings which do not show results for the indices of PV3 and PV4. All the indices are considered together to be included in the discriminant functions, without stepwise method. As the 102 recordings are labelled in four groups, i.e. water, wind, bird and urban, three canonical linear discriminant functions are developed, one less than the number of levels in the group variable. Each function acts as projecting the data onto a dimension that best separates or discriminates between the groups (UCLA Statistical Consulting Group 2013a). The first function provides the most overall discrimination between groups, while the second provides second most, and so on. The functions are independent or orthogonal, that is, their contributions to the discrimination between groups do not overlap (StatSoft Inc. 2013).

For each of the functions, Table 7.4.4 shows the eigenvalue, which indicates the function's discriminating ability, the proportion and cumulative proportion of discriminating ability, and the canonical correlation of predictor variables and the groupings. It can be seen that the first function accounts for 96.0% of the discriminating ability of the discriminating variables or indices, and that the coefficient of canonical correlation is 0.98 of the first function and 0.60 of the second function. These results generally indicate good correlations between the functions and groupings, that is, the functions are effective for the discriminating. Also, Table 7.4.4 shows the tests with the null hypothesis that a function, and all functions that follow, have no discriminating ability. The first test tests all the three functions ("1 through 3"), the second test tests the second and third functions, and the third test tests the third function alone. It shows test results of Wilks' Lambda and Chi-square statistic. The p-value (Sig.) associated with the Chi-square statistic is larger than an alpha level of 0.05 for the third test, which fails to reject the null hypothesis. In other words, the third function does not have much discriminating ability.

Table 7.4.4 Eigenvalues and Wilks' Lambda of discriminant functions based on pitch indices

Eigenvalues					Wilks' Lambda			
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation	Test of Function(s)	Wilks' Lambda	Chi-square	Sig.
1	19.901	96.0	96.0	.976	1 through 3	.024	332.665	.000
2	.552	2.7	98.7	.596	2 through 3	.508	60.604	.006
3	.268	1.3	100.0	.460	3	.788	21.286	.214

Table 7.4.5 Structure matrix of discriminant functions based on pitch indices

	Function 1	Function 2	Function 3
PV AVE	.569*	-.349	-.077
PV STDEV	.530*	-.388	.123
PV STDEVA	.357*	-.319	.053
PV1	.342*	-.157	-.124
PV Percentile25	.319*	-.174	-.074
PV Percentile5	.290*	-.158	-.076
PA AVE	.276*	.264	.048
PV2	.276*	-.106	.209
PV3	.262*	-.123	.102
PV Mode	.231*	-.114	-.054
PV4	.196*	-.022	.185
PA STDEVA	.335	.516*	-.001
PN	-.214	-.455*	.320
PA1	.316	.430*	.267
PA2	.250	.384*	.307
PA3	.236	.288*	.251
PA4	.229	.288*	.209
PA STDEV	.185	.198*	-.006
PV Range	.234	-.393	.696*

Structure matrix of the discriminant functions, shown in Table 7.4.5, displays the correlations between the discriminating variables and the dimensions created with the discriminant functions, in which the variables are ordered by absolute value of correlation within function and * indicates largest absolute correlation between each variable and any discriminant function. The first function has relatively high correlations with pitch values, including Pitch AVE and Pitch STDEV. The second function has relatively high correlations with pitch strengths and the number of pitches, including PA STDEVA, PN and PA1. The third function has relatively high correlation with PV Range. Generally, the first function mainly represents pitch values; the second function mainly represents pitch strengths; and third function mainly represents the range of pitch values.

With the discriminant functions, the 102 recordings are represented in the three-dimensional space created by the three functions, as shown in Figure 7.4.2 (a), where the first and second functions are displayed. Table 7.4.6 shows the functions scores at group centroids, i.e. the means of each of the unstandardized canonical discriminant function scores of recordings in group. The results show that the first discriminant function mainly

discriminates the birdsongs from the other three categories. The second discriminant function mainly discriminates the urban sounds.

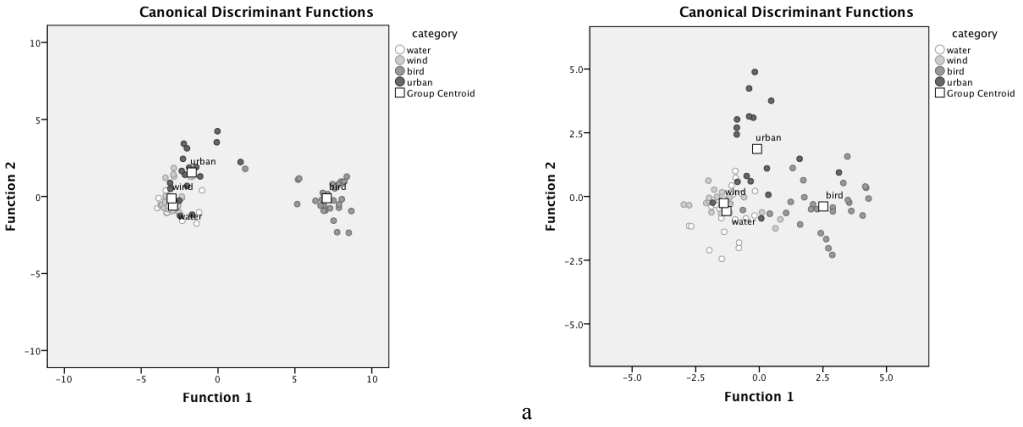
The predicted classification results are shown in Table 7.4.7 in terms of the number and percentage of cases, by both the original functions and cross validations. In cross validation, each case is classified by the functions derived from all cases other than that case. It can be seen that the percentages of correctly classified cases of the water, wind and bird categories are all above 65% for both original functions and cross validations, while of urban category the percentage is somehow lower, about 50% to 60%. Overall for the four categories, 75.5% of originally grouped cases and 73.5% of cross-validated grouped cases are correctly classified. These results suggest the prediction accuracy of pitch indices is generally acceptable, although not high for water, wind and especially urban category.

Table 7.4.6 Group centroids of discriminant functions based on pitch indices and on rhythm indices

Category	Function for Pitch indices			Function for Rhythm indices		
	1	2	3	1	2	3
Water	-2.925	-.597	.477	-1.291	-.572	-.522
Wind	-3.019	-.128	-.869	-1.396	-.262	.851
Bird	7.071	-.116	-.026	2.516	-.391	.043
Urban	-1.711	1.557	.266	-.087	1.866	-.126

Table 7.4.7 Classification results by discriminant functions based on pitch indices and on rhythm indices

		Category	Predicted Group Membership based on Pitch indices					Predicted Group Membership based on Rhythm indices				
			Water	Wind	Bird	Urban	Total	Water	Wind	Bird	Urban	Total
Original	Count	Water	23	9	0	2	34	20	8	0	1	29
		Wind	4	17	0	2	23	5	13	1	0	19
		Bird	0	0	27	1	28	1	1	23	1	26
		Urban	3	4	0	10	17	2	2	1	12	17
	%	Water	**67.6**	26.5	.0	5.9	100.0	**69.0**	27.6	.0	3.4	100.0
		Wind	17.4	**73.9**	.0	8.7	100.0	26.3	**68.4**	5.3	.0	100.0
		Bird	.0	.0	**96.4**	3.6	100.0	3.8	3.8	**88.5**	3.8	100.0
		Urban	17.6	23.5	.0	**58.8**	100.0	11.8	11.8	5.9	**70.6**	100.0
Cross-validated	Count	Water	23	9	0	2	34	14	11	1	3	29
		Wind	5	16	0	2	23	6	10	2	1	19
		Bird	0	0	27	1	28	2	1	22	1	26
		Urban	3	4	1	9	17	4	4	2	7	17
	%	Water	**67.6**	26.5	.0	5.9	100.0	**48.3**	37.9	3.4	10.3	100.0
		Wind	21.7	**69.6**	.0	8.7	100.0	31.6	**52.6**	10.5	5.3	100.0
		Bird	.0	.0	**96.4**	3.6	100.0	7.7	3.8	**84.6**	3.8	100.0
		Urban	17.6	23.5	5.9	**52.9**	100.0	23.5	23.5	11.8	**41.2**	100.0



a b

Figure 7.4.2 Plots of the four categories of sound with discriminant functions based on (a) pitch indices, and (b) rhythm indices

7.4.3.2 Discriminant function analysis based on rhythm indices

DFA is then implemented based on the remaining rhythm indices. All the indices are considered together, i.e., all the indices are included in the discriminant functions. Similarly to that for pitch, three discriminant functions are developed. Table 7.4.8 shows the eigenvalues, the proportion and cumulative proportion of discriminating ability, and the canonical correlation of each given function. The first function accounts for 72.3% of the discriminating ability of the discriminating variables and the second function accounts for 21.4%. The canonical correlation coefficient of the first function is 0.86, and of the second function is 0.69, which generally indicate good correlations between the two functions and groupings, that is, the functions are effective for the discriminating. Table 7.4.8 also shows the tests of functions with the null hypothesis that a function, and all functions that follow, have no discriminating ability, in terms of Wilks' Lambda and Chi-square statistic. The p-values (Sig.) associated with the Chi-square statistic are smaller than an alpha level of 0.05 for the first two tests, and is larger than 0.05 for the third test. That is, for the third test, it fails to reject the null hypothesis. In other words, the third function does not have much discriminating ability.

Table 7.4.8 Eigenvalues and Wilks' Lambda of discriminant functions based on rhythm indices

Eigenvalues					Wilks' Lambda			
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation	Test of Function(s)	Wilks' Lambda	Chi-square	Sig.
1	2.875	72.3	72.3	.861	1 through 3	.111	171.214	.000
2	.850	21.4	93.6	.678	2 through 3	.431	65.567	.004
3	.253	6.4	100.0	.449	3	.798	17.578	.484

Table 7.4.9 Structure matrix of discriminant functions based on rhythm indices

	Function 1	Function 2	Function 3
AS AVE	.735*	.090	.093
AS Percentile90	.569*	.038	-.042
AS Percentile10	.505*	-.010	.220
ED(E)	.457*	.016	.322
BS	.428*	.281	-.031
IOI(E) AVE	-.379*	-.123	-.082
IOI(E) Percentile10	-.350*	-.087	-.011
IOI(E) STDEV	-.316*	-.087	.008
AS STDEV	.282*	.132	-.173
IOI(E) Mode	-.240*	-.043	-.163
SF AVE	.095	.661*	-.036
SF STDEV	.171	.608*	-.062
IOI(SF) Percentile10	.015	.370*	-.216
IOI(SF) Median	.025	.296*	-.203
IOI(SF) Mode	.035	.212*	-.179
ED(SF)	-.243	-.089	.609*
IOI(SF) AVE	.108	.263	-.413*
IOI(SF) STDEV	.187	.143	-.353*
IOI(SF) Percentile90	.142	.095	-.338*
IOI(SF) Range	.208	.134	-.313*

Table 7.4.9 shows the structure matrix of the discriminant functions, the correlations between the discriminating variables and the dimensions created with the discriminant functions. The first function has relatively high correlations with attack slope (AS), event density based on envelope method (ED(E)) and periodicity (BS). The second function has relatively high correlation with spectral flux (SF). The third function has relatively high correlations with event density and event interval based on spectral flux method (ED(SF), IOI(SF)). With the discriminant functions, the 102 recordings are represented in the three-dimensional space created by the three functions, as shown in Figure 7.4.2 (b), where the first and second functions are displayed. Table 7.4.6 shows the functions scores at group centroids. The results show that the first discriminant function mainly discriminates the birdsongs, while the second discriminant function mainly discriminates the urban sounds from the other three categories. Birdsongs have high attack slope, event density based on envelope method and periodicity; urban sounds have high spectral flux.

The predicted classification results are shown in Table 7.4.7. The percentages of correctly classified cases of the four categories are all above 65% for original functions; for cross validations, the percentages are about 40% to 50% for water, wind and bird categories. Overall for the four categories, 74.7% of originally grouped cases and 58.2% of cross-validated grouped cases are correctly classified. Generally, these results suggest the prediction accuracy of rhythm indices is not very high.

7.4.3.3 Discriminant function analysis based on the key pitch and rhythm indices

Additional discriminant functions are explored which based on both key pitch and rhythm indices. The indices are all considered to be included in the functions. Similarly, three functions are developed. From Table 7.4.10, which shows the eigenvalues and the proportion of discriminating ability in a given function, it can be seen that the first function accounts for 89.7% of the discriminating ability of the variables. From the tests of functions with the null hypothesis that they have no discriminating ability, it can be seen that the p-values associated with the Chi-square statistic of all the tests are smaller than an alpha level of 0.05. That is, it rejects the null hypothesis; in other words, all the three functions have the ability for discriminating.

Table 7.4.10 Eigenvalues and Wilks' Lambda of discriminant functions based on key pitch and rhythm indices

Eigenvalues					Wilks' Lambda			
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation	Test of Function(s)	Wilks' Lambda	Chi-square	Sig.
1	10.801	89.7	89.7	.957	1 through 3	.033	317.119	.000
2	.848	7.0	96.8	.677	2 through 3	.390	87.579	.000
3	.388	3.2	100.0	.529	3	.721	30.478	.001

Table 7.4.11 Structure matrix of discriminant functions based on key pitch and rhythm indices

	Function 1	Function 2	Function 3
PV AVE	.770*	.362	.028
PV1	.464*	.180	.088
PA1	.431*	-.319	-.157
PA AVE	.376*	-.177	.002
AS AVE	.341*	-.245	-.145
ED(E)	.240*	-.167	.108
SF AVE	.024	-.577*	.072
BS	.198	-.395*	-.066
ED(SF)	-.103	.077	.485*
IOI(E) AVE	-.151	.195	.375*
PN	-.293	.315	-.330*
IOI(SF) AVE	.032	-.035	-.317*

Structure matrix, as displayed in Table 7.4.11, shows that the first function has relatively high correlation with pitch value and pitch strength; the second function has relatively high correlation with spectral flux; the third function has relatively high correlation with event density based on spectral flux method. These results are somehow similar to those by PCA in Section 7.4.2. Component 1 mainly represents pitch value, pitch strength, percentage of audible pitches over time and event interval, event density and attack slope based on envelope method. Component 2 mainly represents event

interval and event density based on spectral flux method. Component 3 mainly represents spectral flux. In Figure 7.4.3 (a), the 102 recordings are represented in the three-dimensional space created by the three discriminant functions, where the first and second functions are displayed. Table 7.4.12 shows the functions scores at group centroids. It shows that the first discriminant function mainly discriminates the birdsongs, which have high pitch value and strength, from the other three categories; the second discriminant function mainly discriminates the urban sounds, which have high spectral flux; the third discriminant function mainly discriminates between the water and wind sounds, in which wind sounds have higher event density based on spectral flux method.

The predicted classification results displayed in Table 7.4.13 show that the percentages of correctly classified cases of the four categories are all above 65% for original functions and above 55% for cross validations. The percentages of correctly classified cases of the four categories are generally higher than those by the pitch or rhythm indices alone, although a little lower for wind and bird than those by the pitch indices. Overall for the four categories, 78.4% of originally grouped cases and 73.5% of cross-validated grouped cases are correctly classified.

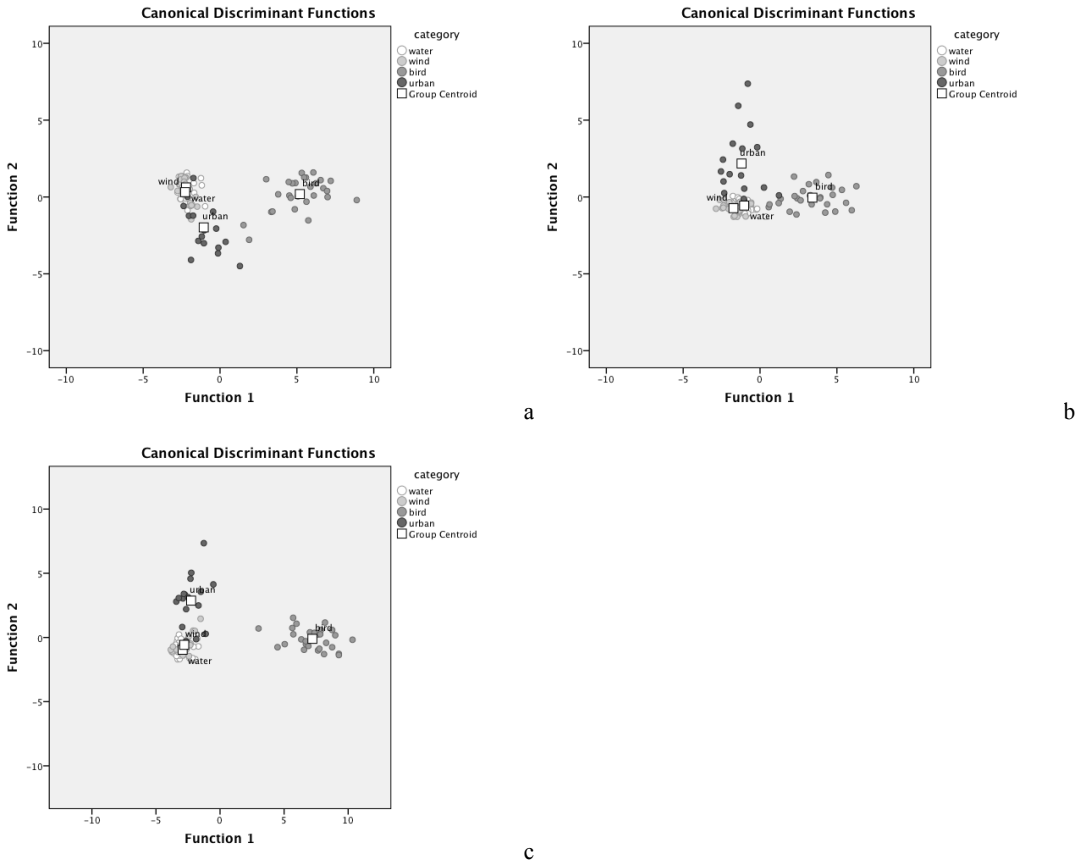


Figure 7.4.3 Plots of the four categories of sound with discriminant functions based on (a) key pitch and rhythm indices, (b) loudness and timbre indices, and (c) all the indices together

Table 7.4.12 Group centroids of discriminant functions based on key pitch and rhythm indices, on loudness and timbre indices, and on all the indices together

Category	Function for Key pitch and rhythm indices			Function for Loudness and timbre indices			Function for Pitch, rhythm, loudness and timbre indices		
	1	2	3	1	2	3	1	2	3
Water	-2.209	.622	-.627	-1.035	-.559	.590	-2.917	-.958	-.922
Wind	-2.271	.314	1.024	-1.730	-.730	-.732	-2.794	-.574	1.523
Bird	5.192	.189	.023	3.407	-.044	-.095	7.208	-.108	-.007
Urban	-1.060	-1.980	-.171	-1.202	2.178	-.033	-2.258	2.870	-.204

Table 7.4.13 Classification results by discriminant functions based on key pitch and rhythm indices, on loudness and timbre indices, and on all the indices together

		Category	Predicted Group Membership based on Pitch and rhythm indices				Predicted Group Membership based on Loudness and timbre indices				Predicted Group Membership based on Pitch, rhythm, loudness, and timbre indices				Total
			Water	Wind	Bird	Urban	Water	Wind	Bird	Urban	Water	Wind	Bird	Urban	
Original	Count	Water	27	6	0	1	28	6	0	0	29	5	0	0	34
		Wind	7	15	0	1	3	20	0	0	2	20	0	1	23
		Bird	0	0	26	2	2	0	26	0	0	0	28	0	28
		Urban	3	2	0	12	3	3	1	10	3	1	0	13	17
	%	Water	**79.4**	17.6	.0	2.9	**82.4**	17.6	.0	.0	**85.3**	14.7	.0	.0	100.0
		Wind	30.4	**65.2**	.0	4.3	13.0	**87.0**	.0	.0	8.7	**87.0**	.0	4.3	100.0
		Bird	.0	.0	**92.9**	7.1	7.1	.0	**92.9**	.0	.0	.0	**100.0**	.0	100.0
		Urban	17.6	11.8	.0	**70.6**	17.6	17.6	5.9	**58.8**	17.6	5.9	.0	**76.5**	100.0
Cross-validated	Count	Water	27	6	0	1	23	11	0	0	28	6	0	0	34
		Wind	9	13	0	1	4	19	0	0	4	18	0	1	23
		Bird	0	0	25	3	3	0	25	0	0	0	27	1	28
		Urban	5	2	0	10	5	3	2	7	4	3	0	10	17
	%	Water	**79.4**	17.6	.0	2.9	**67.6**	32.4	.0	.0	**82.4**	17.6	.0	.0	100.0
		Wind	39.1	**56.5**	.0	4.3	17.4	**82.6**	.0	.0	17.4	**78.3**	.0	4.3	100.0
		Bird	.0	.0	**89.3**	10.7	10.7	.0	**89.3**	.0	.0	.0	**96.4**	3.6	100.0
		Urban	29.4	11.8	.0	**58.8**	29.4	17.6	11.8	**41.2**	23.5	17.6	.0	**58.8**	100.0

7.4.3.4 Discriminant function analysis based on the loudness and timbre indices

In order to examine the contribution of the pitch and rhythm indices to automatic identification of the four categories, and to compare with that of the loudness and timbre indices, additional DFA is implemented based on the loudness and timbre indices. The indices are all considered to be included in the functions. From Table 7.4.14, it can be seen that it can be seen that the p-values associated with the Chi-square statistic of all the tests are smaller than an alpha level of 0.05. That is, it rejects the null hypothesis that they have no discriminating ability; in other words, all the three functions have the ability for discriminating.

In Table 7.4.15, it shows that the first function has relatively high correlations with standard deviation of sharpness and average of fluctuation strength; the second function has relatively high correlations standard deviation and average of tonality; the third

function has certain correlations with average of sharpness and average of roughness. In Figure 7.4.3 (b), where the 102 recordings are represented in the three-dimensional space created by the three discriminant functions, and Table 7.4.12, which shows the functions scores at group centroids, it can be seen that the first discriminant function mainly discriminates the birdsongs from the other three categories; the second discriminant function mainly discriminates the urban sounds; the third discriminant function mainly discriminates between the water and wind sounds.

The predicted classification results displayed in Table 7.4.13 show that the accuracies based on loudness and timbre indices, comparing to those based on pitch and rhythm indices, are higher for water and wind sound categories, equal for birdsong category, and lower for urban sound category. Overall for the four categories, 82.4% of originally grouped cases and 72.5% of cross-validated grouped cases are correctly classified.

Table 7.4.14 Eigenvalues and Wilks' Lambda of discriminant functions based on loudness and timbre indices

Eigenvalues					Wilks' Lambda			
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation	Test of Function(s)	Wilks' Lambda	Chi-square	Sig.
1	4.641	78.0	78.0	.907	1 through 3	.069	251.349	.000
2	1.057	17.8	95.8	.717	2 through 3	.389	88.726	.000
3	.249	4.2	100.0	.447	3	.800	20.928	.007

Table 7.4.15 Structure matrix of discriminant functions based on loudness and timbre indices

	Function 1	Function 2	Function 3
S STDEV	.667*	.134	-.261
Fls AVE	.445*	.441	-.167
Ton STDEV	.108	.736*	-.232
Ton AVE	.058	.677*	-.135
Fls STDEV	.398	.425*	-.193
N STDEV	.011	.333*	-.116
R STDEV	.092	.286*	.049
N AVE	-.231	.239*	-.072
S AVE	.304	.060	.321*
R AVE	-.239	.102	.251*

7.4.3.5 Discriminant function analysis based on the pitch, rhythm, loudness and timbre indices

All the pitch and rhythm indices the loudness and timbre indices are all considered together to be included in the discriminant functions. From Table 7.4.16, it can be seen that it can be seen that the p-values associated with the Chi-square statistic of all the tests are smaller than an alpha level of 0.05. That is, it rejects the null hypothesis that they

have no discriminating ability; in other words, all the three functions have the ability for discriminating.

Table 7.4.17 shows that the first function has relatively high correlations with pitch values; the second function has relatively high correlations with tonality; the third function has relatively certain correlations with event density based on spectral flux method and sharpness. In Figure 7.4.3 (c), where the 102 recordings are represented in the three-dimensional space created by the three discriminant functions, and Table 7.4.12, which shows the functions scores at group centroids, it can be seen that the first discriminant function mainly discriminates the birdsongs from the other three categories; the second discriminant function mainly discriminates the urban sounds; the third discriminant function mainly discriminates between the water and wind sounds.

Table 7.4.16 Eigenvalues and Wilks' Lambda of discriminant functions based on pitch, rhythm, loudness, and timbre indices

Eigenvalues					Wilks' Lambda			
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation	Test of Function(s)	Wilks' Lambda	Chi-square	Sig.
1	20.511	88.5	88.5	.976	1 through 3	.009	410.753	.000
2	1.828	7.9	96.3	.804	2 through 3	.192	143.790	.000
3	.847	3.7	100.0	.677	3	.542	53.357	.000

Table 7.4.17 Structure matrix of discriminant functions based on pitch, rhythm, loudness, and timbre indices

	Function 1	Function 2	Function 3
PV AVE	.563*	-.097	-.060
PV1	.338*	-.034	.013
S STDEV	.319*	.070	-.085
PA1	.306*	.301	-.139
PA AVE	.269*	.193	-.030
AS AVE	.242*	.233	-.124
ED(E)	.171*	.159	.054
R AVE	-.116*	.078	-.081
Ton STDEV	.058	.558*	-.048
Ton AVE	.033	.512*	-.073
SF AVE	.008	.396*	.059
Fls AVE	.215	.310*	-.127
BS	.137	.307*	-.055
Fls STDEV	.193	.303*	-.096
L AVE	-.096	.277*	.173
PN	-.208	-.268*	-.203
N STDEV	.008	.256*	-.004
R STDEV	.044	.205*	-.107
N AVE	-.107	.199*	.062
ED(SF)	-.072	-.077	.336*
S AVE	.139	.004	-.272*
IOI(E) AVE	-.105	-.165	.264*
IOI(SF) AVE	.022	.033	-.217*
L STDEV	.121	.094	-.148*

The predicted classification results displayed in Table 7.4.13 show that when all the loudness, timbre, pitch and rhythm indices are used together, the proportions of correctly classified cases are all above 75% for the four sound categories with original functions and for the three natural sound categories in cross validations, and is 59% for the urban sound category in cross validation. The accuracies are higher than either those based on pitch and rhythm indices, or those based on loudness and timbre indices for all the four categories. Overall for the four categories, 88.2% of originally grouped cases and 81.4% of cross-validated grouped cases are correctly classified. These results suggest the prediction accuracy of all the indices is generally good, except for urban category. Also, these results indicate the contribution of the pitch and rhythm indices to the automatic identification of environmental sound type.

7.5 Conclusions

Based on the algorithms of pitch and rhythm features selected for environmental sound in Chapters 5 and 6 and a number of parameters and statistic indices developed from these algorithms, the correlations between the pitch and rhythm indices and those psychoacoustic ones that are used in Chapter 4, i.e. loudness and timbre indices, are examined. Generally the correlations between the two sets of indices are not very high (coefficients generally below 0.8), although there are certain correlations between (coefficients of about 0.6 to 0.8), e.g., pitch and sharpness, pitch strength and tonality, attack slope and fluctuation, and spectral flux and SPL or loudness. It suggests that generally the two sets of indices do not share much common variances. Thus, the pitch and rhythm indices developed in Chapters 5 and 6 provide additional variance to the previous psychoacoustic indices that have been analysed with in Chapter 4.

For both pitch and rhythm, the PCAs suggest that the statistic indices of each parameter, such as average, standard deviation, medium and percentiles, are mostly in one single dimension, or say contribute to one component. Thus, among the indices, based on the PCAs and correlations between the indices, a number of key indices are identified, which generally contribute most to each of the PCs. For pitch, they are pitch value and pitch strength of the most distinct pitch over the whole duration (PV1, PA1), averages of pitch values and pitch strengths over time (PV AVE, PA AVE), and percentage of audible pitches over time (PN). For rhythm, they are average event interval, event density and average attack slope of event based on envelope method (IOI(E) AVE, ED(E), AS AVE), average event interval, event density and average spectral flux based on spectral flux method (IOI(SF) AVE, ED(SF), SF AVE), and periodicity (BS).

With these indices, the different characteristics of different environmental sounds are shown. Generally, in terms of pitch, water sounds have low pitch values, low pitch strengths and high percentage of audible pitches; wind sounds have low pitch values and low pitch strengths; birdsongs have high pitch values, high pitch strengths and low percentage of audible pitches; and urban sounds have low pitch values and a relatively wide range of pitch strengths. In terms of rhythm, water and wind sounds both have relatively high event interval, low event density, and low attack slope of event based on the envelope method; oppositely, birdsongs have relatively low event interval, high event density, and high attack slope of event; and urban sounds have a relatively wide range of event interval, event density and attack slope. A number of birdsongs and urban sounds show periodicity, while water and wind sounds do not. The indices based on spectral flux method generally do not show any significant differences among.

Moreover, the pitch and rhythm indices contribute to the automatic identification of environmental sound type. When all the loudness, timbre, pitch and rhythm indices are used together, the proportions of correctly classified cases are above 85% for the three natural sound categories and above 75% for the urban sound category, all of which are higher than those based on loudness and timbre indices only.

Chapter 8
1/f noise behaviour of natural and urban sounds in soundscapes

The 1/f noise behaviour can be seen as a statistic index in addition to those commonly used, such as mean, standard deviation, maximum and minimum in Chapters 4 and 7. In this chapter, the 1/f noise behaviours of four different categories of sound in the loudness, timbre and pitch parameters in those two chapters are analysed in Section 8.1. In addition, in Section 8.2, it explores the 1/f noise of signal in each critical band related to the auditory system, based on the specific loudness.

8.1 1/f Noise Behaviours of Environmental Sounds in Terms of Psychoacoustic and Music Parameters

The 1/f noise behaviours of four different categories of sound, i.e. water, wind, birdsong and urban sounds, are analysed, based on the results of the psychoacoustic and music parameters in Chapters 4 and 7. These parameters include loudness, sharpness, tonality and pitch. As 1/f noise measures dynamic of sound in terms of a given parameter, the parameters that reflect the variation of sound with time are not included, e.g. roughness, fluctuation strength, and the rhythm parameters.

8.1.1 1/f noise behaviours of loudness

8.1.1.1 Frequency range of spectrum density

As described in Chapter 3, the duration of the recordings used in the analysis vary among 30s, 60s, 120s and 240s, according to the availability. Since the majority of the recordings are of duration of 240s, 1/f noise behaviours of the 102 recordings are analysed over the frequency range of 0.005-10Hz, which responds to the time range of 0.1-200s. It is noted, consequently, that some of the recordings may not be analysed for the full frequency range because of the limitation of duration; however, it may not effect the main results to a large degree.

By exploring the spectrum density of variation of loudness with time for the 102 recordings, it can be seen that while typical 1/f noise exhibit a slope of spectrum density of -1, the slopes of the environmental sounds vary. Figure 8.1.1 shows several types of the shape of spectrum density of loudness with time, in which (a) - (e) respectively respond to the recordings of a river, a sea waves, two winds and a birdsong. It shows that the shape of spectrum density may not be a straight line, but with breaks at some points. For example, in Figure 8.1.1 (b) there is a break in the curve at the place corresponding to between -1.0 and -0.5 on abscissa; in Figure 8.1.1 (c) a break occurs at about 0.0 on abscissa; in Figure 8.1.1 (e) breaks occur at about -1.0 and 0.0 on abscissa; and in Figure 8.1.1 (a) and (d) the shapes of spectrum density are generally in a line, through with different slopes.

In order to describe more precisely the shape and slope of spectrum density, it may be useful to further divide the analysed frequency range into several sub-ranges. Since by examining the shapes of spectrum density of the 102 recordings, breaks often occur at points corresponding to about -1.0 and 0.0 on abscissa, the full frequency range of [0.005-10Hz] is divided into three ranges, i.e., [0.005-0.1Hz], [0.1-1Hz] and [1-10Hz], which respond to [-2.3, -1.0], [-1.0, 0.0] and [0.0, 1.0] in logarithmic scale in the figures, and respond to the time ranges of [10-200s], [1-10s] and [0.1-1s] respectively. Consequently, in each of the frequency ranges, the slope of the spectrum density and its corresponding deviation are calculated. That is, 8 indices are derived from the spectrum density, which are slope and deviation of the frequency ranges of [0.005-0.1Hz], [0.1-1Hz], [1-10Hz], and [0.005-10Hz] which as a whole.

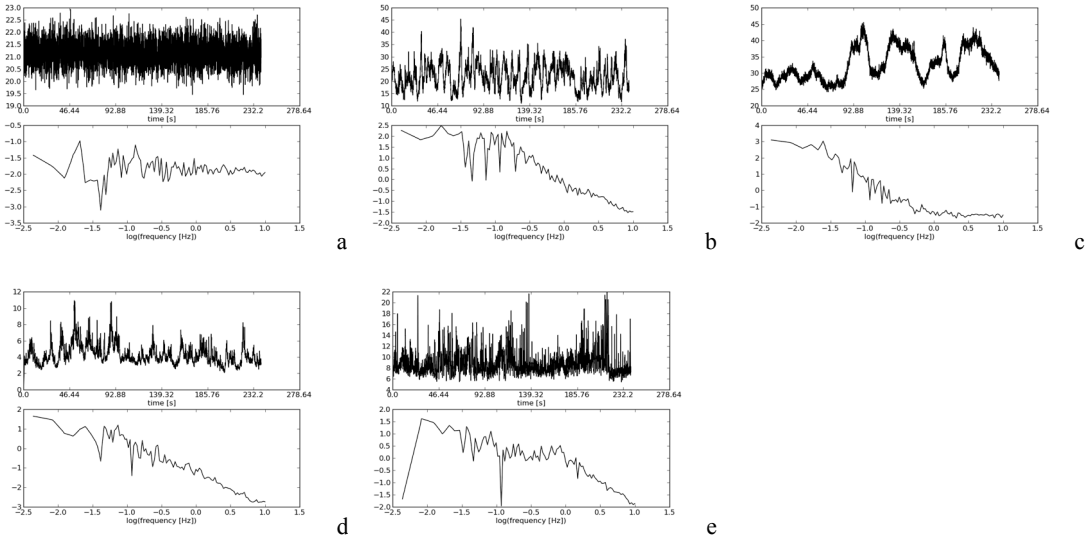


Figure 8.1.1 Shapes of spectrum density of loudness, illustrated by the No. 1,17, 37, 53, and 59 cases of the recordings

8.1.1.2 Comparison among the categories by the means of 1/f noise indices of loudness with one-way analysis of variance

The means in these 8 indices among the four categories are compared with one-way analysis of variance (ANOVA) in the software of SPSS Statistics. The descriptives of the indices for the four categories are shown in Table 8.1.1, including mean, standard deviation, minimum, and maximum. Table 8.1.2 shows the results of ANOVA, in terms of the F ratio and p value, i.e. the significance of the F ratio (Sig.). The F ratio shows the ratio of between-groups variance to within-groups variance. For the ones that p value is less than or equal to an alpha level of 0.05, the corresponding null hypothesis that the means are equal are rejected, while for the ones that p value is larger than 0.05, it is failed to reject the null hypothesis, meaning that it is unlikely that these means differ. From the table, it can be seen that the p value of N deviation [0.005-10Hz] (deviation of the full frequency range) is greater than the α level of 0.05, which suggests that there may not be statistically significantly different among the four categories in the means of the index; for the other indices, the p value associated with the F ratio is less than 0.05, which reject the null hypothesis that all the means are equal. In other words, except for deviation of the full frequency range, all the other indices show some significant differences among the categories, or between at least two categories, whereas the assumption of homogeneity of variances and post hoc tests are further checked for verifying and examining which of the specific categories differ.

The test of homogeneity of variances is examined to test the assumption of ANOVA that all variances of different categories are equal. Table 8.1.2 shows the results of Levene's test of homogeneity of variances of the four categories in the 1/f noise indices of loudness indices. It shows that p value (Sig.) of N deviation [0.005-0.1Hz] is greater than an α level of 0.05, which fails to reject the null hypothesis, meaning that the variances are equal and the assumption of homogeneity of variance is met. For other indices, the p values (Sig.) are less than α level of 0.05, thus the assumptions are rejected, suggesting the variances are unequal.

As differences exist among the means of the categories in the indices except deviation of the full frequency range, post hoc tests are executed to examine which means significantly differ from the others. A number of different methods are available to produce post hoc tests based on the assumption of equal variances or not. Here, for the indices with equal variances between categories, among a variety of most common tests, including least significant difference (LSD) test developed by Fisher, Tukey's honestly significant difference (Tukey's HSD) test, the Student-Newman-Keuls (SNK) test and Scheffé's test, Tukey's HSD test is used, considering the unequal sample sizes in groups (Stevens 1999). For situations that variances are unequal and sample sizes are unequal,

Dunnett's T3 method is used among the available methods, including Tamhane's T2, Dunnett's T3, and Games - Howell and Dunnett's C.

Table 8.1.1 Descriptives of 1/f noise indices of loudness for the four categories

		N	Mean	Std. Deviation	Minimum	Maximum		N	Mean	Std. Deviation	Minimum	Maximum
Water	N	34	-0.850	0.589	-1.587	0.074	N	34	0.479	0.140	0.269	0.738
Wind	slope	23	-1.638	0.123	-1.878	-1.335	deviation	23	0.461	0.079	0.314	0.582
Bird	0.005Hz	28	-0.801	0.160	-1.247	-0.517	0.005Hz	28	0.456	0.074	0.348	0.631
Urban	10Hz	17	-1.094	0.465	-2.111	-0.589	10Hz	17	0.467	0.112	0.344	0.714
Water	N	34	-0.805	0.570	-2.023	0.134	N	34	0.085	0.015	0.059	0.125
Wind	slope	23	-0.495	0.401	-1.770	-0.053	deviation	23	0.076	0.011	0.064	0.106
Bird	1Hz	28	-1.133	0.409	-1.844	-0.349	1Hz	28	0.113	0.029	0.074	0.169
Urban	10Hz	17	-1.055	0.549	-1.808	-0.211	10Hz	17	0.192	0.067	0.098	0.343
Water	N	34	-1.434	1.313	-3.791	0.237	N	34	0.234	0.033	0.170	0.312
Wind	slope	23	-2.052	0.370	-2.840	-1.325	deviation	23	0.212	0.032	0.136	0.264
Bird	0.1Hz	28	-1.210	0.616	-2.253	0.229	0.1Hz	28	0.243	0.030	0.179	0.304
Urban	1Hz	17	-1.108	0.799	-2.922	-0.006	1Hz	17	0.308	0.090	0.187	0.532
Water	N	34	-0.336	0.459	-1.405	0.548	N	34	0.231	0.063	0.079	0.395
Wind	slope	23	-1.519	0.446	-2.158	-0.745	deviation	23	0.242	0.060	0.163	0.354
Bird	0.005Hz	28	-0.386	0.499	-1.275	0.518	0.005Hz	28	0.244	0.056	0.151	0.368
Urban	0.1Hz	17	-0.839	0.818	-2.194	0.983	0.1Hz	17	0.137	0.067	0.011	0.286

Table 8.1.2 Test of homogeneity of variances and ANOVA of 1/f noise indices of loudness for the four categories

	Test of Homogeneity of Variances		ANOVA			Test of Homogeneity of Variances		ANOVA	
	Levene Statistic	Sig.	F	Sig.		Levene Statistic	Sig.	F	Sig.
N slope [0.005-10Hz]	48.379	0	22.717	0	N deviation [0.005-10Hz]	8.502	0	0.255	0.858
N slope [1-10Hz]	3.642	0.015	8.150	0	N deviation [1-10Hz]	21.15	0	50.265	0
N slope [0.1-1Hz]	27.158	0	4.853	0.003	N deviation [0.1-1Hz]	16.486	0	14.871	0
N slope [0.005-0.1Hz]	2.868	0.040	26.180	0	N deviation [0.005-0.1Hz]	0.088	0.966	13.341	0

The results of the post hoc tests are shown in Table 8.1.3, by either Tukey's HSD or Dunnett's T3 method according to the homogeneity of variances of the categories in the index. It shows the pairwise multiple comparisons among the categories, in terms of difference between the means of each category with each of the other three, where asterisks indicate significantly different group means at an alpha level of 0.05. From the table, it can be seen that the slope of spectrum density of loudness over the full range of [0.005-10Hz] (N slope [0.005-10Hz]) and the slope over the range of [0.005Hz-0.1Hz] (N slope [0.005-0.1Hz]) show significant differences between wind and the other three categories, while the slope over the range of [1Hz-10Hz] (N slope [1-10Hz]) and the slope over the range of [0.1Hz-1Hz] (N slope [0.1-1Hz]) show significant differences

between the categories of wind and bird, and between wind and urban. Table 8.1.1 shows that for N slope [0.005-10Hz], the mean value is -1.6 for wind category, and is -0.8 to -1.1 for the other three categories. For N slope [0.005-0.1Hz], the mean value of wind category is -1.5, of both water and bird categories are -0.3 to -0.4, and of urban category is -0.8. For N slope [1-10Hz], the mean value(s) of wind category is -0.5, and of water, bird and urban categories are -0.8 to -1.1. For N slope [0.1-1Hz], the mean value(s) of wind category is -2.1, and of the other three categories are -1.1 to -1.4. These results suggest that the means of slopes of spectrum density of loudness over the different ranges mainly distinguish wind sounds from the other three categories of sounds.

In terms of deviations from the slopes of spectrum density of loudness, deviation over the range of [1Hz-10Hz] shows significant differences between each pair of categories. The mean value of deviation of urban category is higher than that of bird category, than water category, and than wind category. Deviation over the range of [0.1Hz-1Hz] shows the significant differences between water and urban, wind and bird, and wind and urban. Urban category has higher mean value than the other three categories. Deviation over the range of [0.005Hz-0.1Hz] shows the significant differences between urban sounds and the other three categories. Urban category has lower mean value than the other three categories. Deviation of the full frequency range [0.005-10Hz] does not show any significant differences among the categories.

Table 8.1.3 Multiple Comparisons of 1/f noise indices of loudness for the four categories

Mean Difference (I-J)	(I) Category	Water			Wind			Bird			Urban		
Dependent Variable	(J) Category	Wind	Bird	Urban	Water	Bird	Urban	Water	Wind	Urban	Water	Wind	Bird
N slope [0.005-10Hz]	Dunnett T3	0.788*	-0.049	0.244	-0.788*	-0.837*	-0.544*	0.049	0.837*	0.292	-0.244	0.544*	-0.292
N deviation [0.005-10Hz]	Dunnett T3	0.018	0.023	0.012	-0.018	0.005	-0.006	-0.023	-0.005	-0.010	-0.012	0.006	0.010
N slope [1-10Hz]	Dunnett T3	-0.310	0.328	0.250	0.310	0.638*	0.560*	-0.328	-0.638*	-0.078	-0.250	-0.560*	0.078
N deviation [1-10Hz]	Dunnett T3	0.009*	-0.028*	-0.106*	-0.009*	-0.038*	-0.116*	0.028*	0.038*	-0.078*	0.106*	0.116*	0.078*
N slope [0.1-1Hz]	Dunnett T3	0.619	-0.224	-0.325	-0.619	-0.842*	-0.944*	0.224	0.842*	-0.102	0.325	0.944*	0.102
N deviation [0.1-1Hz]	Dunnett T3	0.022	-0.009	-0.074*	-0.022	-0.031*	-0.096*	0.009	0.031*	-0.065	0.074*	0.096*	0.065
N slope [0.005-0.1Hz]	Dunnett T3	1.184*	0.050	0.504	-1.184*	-1.133*	-0.680*	-0.050	1.133*	0.453	-0.504	0.680*	-0.453
N deviation [0.005-0.1Hz]	Tukey HSD	-0.010	-0.013	0.095*	0.010	-0.003	0.105*	0.013	0.003	0.108*	-0.095*	-0.105*	-0.108*

8.1.1.3 Characteristics of the categories of sound in terms of 1/f noise indices of loudness

These characteristics of the four categories of sounds in terms of the indices can also be seen in Figure 8.1.2 (a-d), where the 102 sound recordings are plotted in the two-

dimensional coordinate systems with their axes presenting respectively the indices of slope and deviation in the different ranges. It can be seen from the plots, especially in Figure 8.1.2 (a) and (c), that the recordings in water sound category gather in two groups. Detailed data show that the recordings of stream and river sounds in water category are in one of the groups, while the recordings of sea waves sound are in the other. Specifically, in the full range of [0.005-10Hz], both sea waves sounds and wind sounds have mean slope values of about -1.5, stream and river sounds have mean slope values of about -0.5 to 0.0, and birdsongs and urban sounds have mean slope values close to -1.0. The deviations of sea waves sounds are higher than those of wind, birdsongs and urban sounds, and than stream and river sounds. In the range of [1Hz-10Hz], wind sounds have mean slope value of about -0.5, while water, birdsongs and urban sounds have slope values spreading between about -2.0 and 0.0. The mean value of deviation of urban sounds is higher than that of birdsongs, than water sounds, and than wind sounds. In the range of [0.1Hz-1Hz], stream and river sounds have mean slope values of about 0.0, sea waves sounds and wind sounds have mean slope values of about or lower than -2.0, and birdsongs and urban sounds have mean slope values close to -1.0. Urban sounds have higher mean deviation value than the other three categories of sounds. In the range of [0.005Hz-0.1Hz], wind sounds have mean slope value of -1.5, both water sounds and birdsongs have mean slope values of about -0.5 to 0.0, and urban sounds have mean slope value of about -1.0. Urban sounds have lower mean deviation value than the other three categories of sounds.

These characteristics of the four categories can again be seen in Figure 8.1.2 (e) and (f), where the 102 sound recordings are plotted in the three-dimensional coordinate systems with their axes presenting the slope indices in the three frequency ranges. From the figures and the results above, it can be seen that stream and river sounds have mean slope values of about 0.0 in all the three frequency ranges. Sea waves sounds have mean slope values of about -1.0 in the range of [1Hz-10Hz], about -2.0 in the range of [0.1Hz-1Hz], and about 0.0 in the range of [0.005Hz-0.1Hz]. Wind sounds have mean slope values of about -0.5 in the range of [1Hz-10Hz], about -2.0 in the range of [0.1Hz-1Hz], and about -1.5 in the range of [0.005Hz-0.1Hz]. Birdsongs have mean slope values about -1.0 in the ranges of [1Hz-10Hz] and [0.1Hz-1Hz], and about -0.5 to 0.0 in the range of [0.005Hz-0.1Hz]. While the recordings in water, wind and bird categories are generally apart from each others, the recordings in urban category are more dispersive, mixed with many of the recordings in the other three categories.

In other words, stream and river sounds in water category exhibit quick variations in loudness in the three frequency ranges, which respond to 0.1-1s, 1-10s, and 10-200s. Sea waves sounds in water category exhibit generally 1/f noise behaviour in loudness in short

range of 1-10s, slow variation in medium time range of 0.1-1s, and quick variation in long time range of 10-200s. Wind sounds exhibit quick variation in loudness are in short time range of 0.1-1s, and slow variation in medium and long time ranges of 1-10s and 10-200s. Birdsongs exhibit generally 1/f noise behaviour in short and medium time ranges, and quick variation in loudness in long time range.

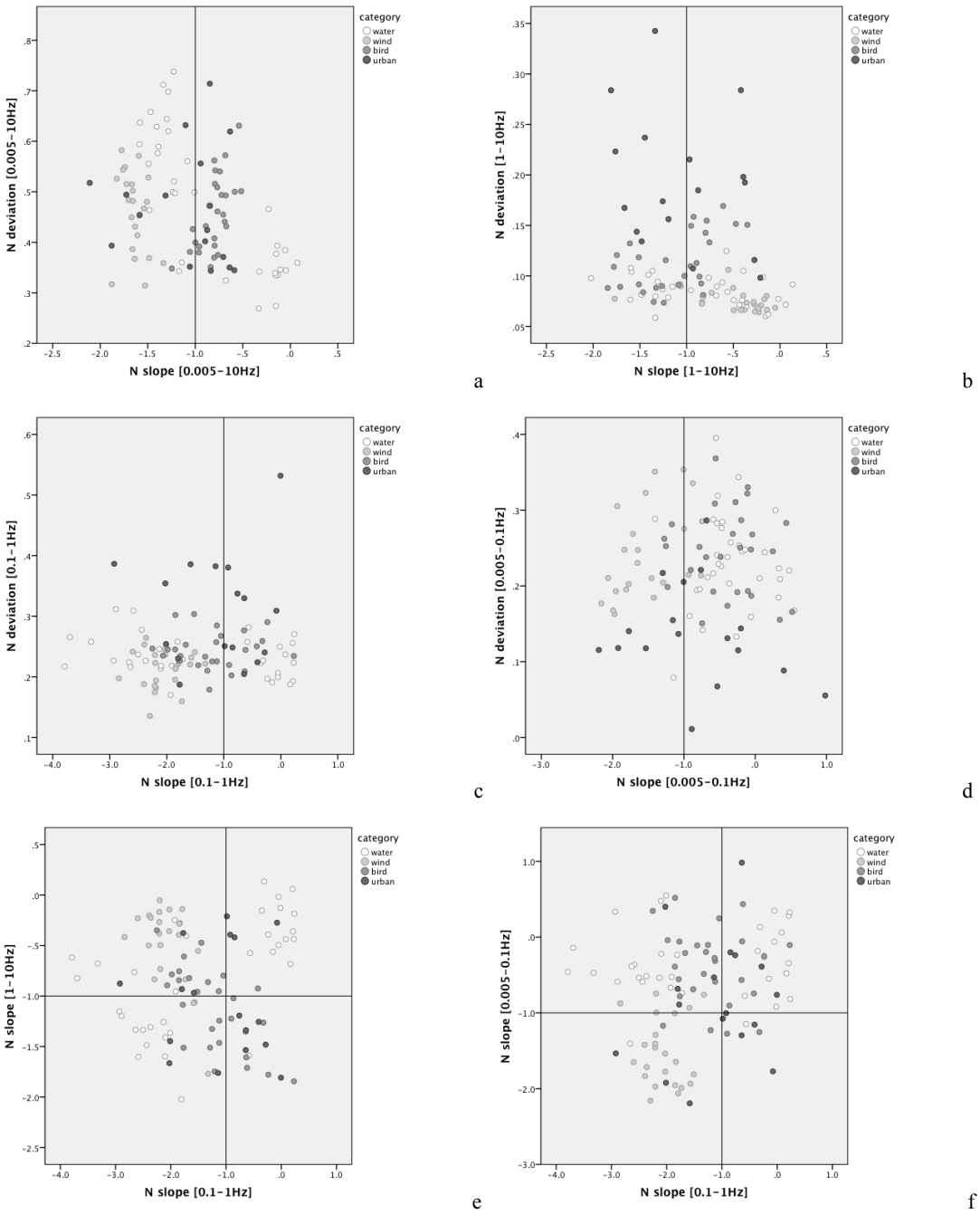


Figure 8.1.2 Characteristics of the four categories of sound in terms of 1/f noise of loudness

8.1.2 1/f noise behaviours of sharpness

The same as that of loudness, 1/f noise behaviours of the 102 recordings in sharpness are analysed over the frequency range of 0.005-10Hz. By examining the shapes of the spectrum density of variation of sharpness with time of the recordings, it is found that the shape types are somehow similar to those of loudness. Thus, the same as that for loudness, the analysed frequency range is divided into three ranges, i.e., [0.005-0.1Hz], [0.1-1Hz] and [1-10Hz], which respond to time ranges of [10-200s], [1-10s] and [0.1-1s] respectively. Consequently, 8 indices are derived from the spectrum density, i.e., slope and deviation of frequency ranges of [0.005-0.1Hz], [0.1-1Hz], [1-10Hz], and [0.005-10Hz] which as a whole.

8.1.2.1 Comparison among the categories by the means of 1/f noise indices of sharpness with one-way analysis of variance

Table 8.1.4 Descriptives of 1/f noise indices of sharpness for the four categories

		N	Mean	Std. Deviation	Minimum	Maximum		N	Mean	Std. Deviation	Minimum	Maximum
Water	S	34	-0.698	0.474	-1.306	0.019	S	34	0.492	0.122	0.336	0.738
Wind	slope	23	-1.414	0.259	-1.657	-0.672	deviation	23	0.430	0.057	0.318	0.531
Bird	[0.005-	28	-0.852	0.161	-1.195	-0.580	[0.005-	28	0.450	0.078	0.312	0.633
Urban	10Hz]	17	-0.919	0.404	-1.708	-0.321	10Hz]	17	0.424	0.076	0.332	0.550
Water	S	34	-0.643	0.452	-1.519	-0.021	S	34	0.082	0.023	0.060	0.195
Wind	slope	23	-0.633	0.247	-1.290	-0.317	deviation	23	0.078	0.021	0.065	0.163
Bird	[1-	28	-1.007	0.409	-1.671	-0.258	[1-	28	0.106	0.029	0.066	0.171
Urban	10Hz]	17	-0.872	0.508	-1.451	0.217	10Hz]	17	0.187	0.075	0.078	0.334
Water	S	34	-1.275	1.127	-2.979	0.287	S	34	0.246	0.034	0.183	0.326
Wind	slope	23	-1.613	0.547	-2.477	-0.364	deviation	23	0.217	0.021	0.175	0.264
Bird	[0.1-	28	-1.374	0.486	-2.163	-0.042	[0.1-	28	0.235	0.029	0.192	0.308
Urban	1Hz]	17	-1.069	0.590	-2.726	-0.216	1Hz]	17	0.277	0.068	0.192	0.427
Water	S	34	-0.213	0.657	-2.308	1.368	S	34	0.269	0.065	0.151	0.450
Wind	slope	23	-1.424	0.472	-2.391	-0.371	deviation	23	0.249	0.068	0.114	0.387
Bird	[0.005-	28	-0.387	0.497	-1.368	0.535	[0.005-	28	0.251	0.060	0.148	0.359
Urban	0.1Hz]	17	-0.952	0.730	-2.096	0.333	0.1Hz]	17	0.145	0.079	0.003	0.287

Table 8.1.5 Test of homogeneity of variances and ANOVA of 1/f noise indices of sharpness for the four categories

	Test of Homogeneity of Variances		ANOVA			Test of Homogeneity of Variances		ANOVA	
	Levene Statistic	Sig.	F	Sig.		Levene Statistic	Sig.	F	Sig.
S slope [0.005-10Hz]	22.524	0	19.780	0	S deviation [0.005-10Hz]	6.808	0	3.171	0.028
S slope [1-10Hz]	6.492	0	5.328	0.002	S deviation [1-10Hz]	23.791	0	34.289	0
S slope [0.1-1Hz]	21.143	0	1.698	0.172	S deviation [0.1-1Hz]	9.043	0	8.461	0
S slope [0.005-0.1Hz]	1.002	0.395	22.615	0	S deviation [0.005-0.1Hz]	0.560	0.642	13.824	0

Table 8.1.6 Multiple Comparisons of 1/f noise indices of sharpness for the four categories

Mean Difference (I-J)	(I) Category	Water			Wind			Bird			Urban		
Dependent Variable	(J) Category	Wind	Bird	Urban	Water	Bird	Urban	Water	Wind	Urban	Water	Wind	Bird
S slope [0.005-10Hz]	Dunnett T3	0.717*	0.155	0.221	-0.717*	-0.562*	-0.495*	-0.155	0.562*	0.067	-0.221	0.495*	-0.067
S deviation [0.005-10Hz]	Dunnett T3	0.062	0.042	0.069	-0.062	-0.020	0.006	-0.042	0.020	0.026	-0.069	-0.006	-0.026
S slope [1-10Hz]	Dunnett T3	-0.010	0.365*	0.230	0.010	0.374*	0.239	-0.365*	-0.374*	-0.135	-0.230	-0.239	0.135
S deviation [1-10Hz]	Dunnett T3	0.004	-0.024*	-0.105*	-0.004	-0.027*	-0.109*	0.024*	0.027*	-0.081*	0.105*	0.109*	0.081*
S slope [0.1-1Hz]	Dunnett T3	0.338	0.099	-0.206	-0.338	-0.240	-0.544*	-0.099	0.240	-0.305	0.206	0.544*	0.305
S deviation [0.1-1Hz]	Dunnett T3	0.030*	0.011	-0.030	-0.030*	-0.019	-0.060*	-0.011	0.019	-0.041	0.030	0.060*	0.041
S slope [0.005-0.1Hz]	Tukey HSD	1.212*	0.175	0.740*	-1.212*	-1.037*	-0.472	-0.175	1.037*	0.565*	-0.740*	0.472	-0.565*
S deviation [0.005-0.1Hz]	Tukey HSD	0.020	0.018	0.124*	-0.020	-0.002	0.103*	-0.018	0.002	0.105*	-0.124*	-0.103*	-0.105*

In order to examine if the sound categories differ from each other significantly in the indices, the means among the four categories are compared with ANOVA, in terms of these 8 indices. The descriptives of the 1/f noise indices of sharpness for the four categories are shown in Table 8.1.4, including mean, standard deviation, minimum, and maximum. Before the ANOVA, the homogeneity of variances of different categories, as the assumption of ANOVA, is firstly tested, the results of which are shown in Table 8.1.5. It shows that p value (Sig.) of S slope [0.005-0.1Hz] and S deviation [0.005-0.1Hz] are greater than an α level of 0.05, which means it fails to reject the null hypothesis that the variances of the categories are equal, that is, the assumption of homogeneity of variance has been met. For the other indices, the p values are less than the α level of 0.05, then the hypothesis that the variances are equal is rejected; in other words, the variances are unequal. Table 8.1.5 also shows the results of ANOVA, in terms of F ratio and p value, i.e. the significance of the F ratio (Sig.). The p value of S slope [0.1-1Hz] is greater than an α level of 0.05, which fails to reject the null hypothesis that all the means are equal; for the other indices, the p values are less than 0.05, which reject the null hypothesis. That is, except for S slope [0.1-1Hz], all the indices show significant differences among the means of the categories, or between at least two categories. However, since the variances are unequal, the results may be inaccurate, and thus further post hoc tests are checked for verifying and examining which of the specific categories differ.

The results of the post hoc tests are shown in Table 8.1.6, by either Tukey's HSD or Dunnett's T3 method according to the homogeneity of variances of the categories in the index. It shows the pairwise multiple comparisons among the categories, in terms of difference between the means of each category with each of the other three, where asterisks indicate significantly different group means at an alpha level of 0.05. As shown

in the table, the slope of spectrum density of sharpness over the full range [0.005-10Hz] (S slope [0.005-10Hz]) shows significant differences between wind and the other three sound categories. In Table 8.1.4, it shows that the mean value of S slope [0.005-10Hz] is -1.4 for wind category, and is -0.7 to -0.9 for the other three categories. The slope over the range of [1Hz-10Hz] (S slope [1-10Hz]) shows significant differences between the categories of bird and wind, and between bird and water. The mean value of S slope [1-10Hz] is -0.6 for both water and wind categories, -1.0 for bird category, and -0.9 for urban category. The slope over the range of [0.1Hz-1Hz] (S slope [0.1-1Hz]) shows significant differences between the categories of wind and urban. The mean value of S slope [0.1-1Hz] is -1.6 for wind categories, -1.1 for urban category, and -1.3 to -1.4 for water and bird categories. The slope over the range of [0.005Hz-0.1Hz] shows the significant differences between the categories of water and bird and the categories of wind and urban. The mean value of S slope [0.005Hz-0.1Hz] is -0.2 to -0.4 for water and bird categories, and is -1.0 to -1.4 for wind and urban categories.

In terms of deviations from the slopes of spectrum density of sharpness, deviation over the range of [1Hz-10Hz] shows significant differences between each pair of categories except for water and wind. The mean value of deviation of urban category is higher than that of bird category, and than water and wind. Deviation over the range of [0.1Hz-1Hz] shows the significant differences between wind and water, and between wind and urban. Urban category has higher mean value than water and bird categories, and than wind category. Deviation over the range of [0.005Hz-0.1Hz] shows the significant differences between urban sounds and the other three categories. Urban category has lower mean value than the other three categories. Deviation of the full frequency range [0.005-10Hz] does not show any significant differences among the categories.

8.1.2.2 Characteristics of the categories of sound in terms of 1/f noise indices of sharpness

These characteristics of the four categories of sounds in terms of the indices can also be seen in Figure 8.1.3 (a-d), where the 102 sound recordings are plotted in the two-dimensional coordinate systems with their axes presenting respectively the indices of slope and deviation in the different ranges. Similarly to loudness in Section 8.1.1, the recordings in water sound category gather in two groups in the plots, especially in Figure 8.1.2 (a) and (c). These two groups respectively are stream and river sound recordings and sea waves sound recordings. In the full range of [0.005-10Hz], wind sounds have mean slope values of about -1.5; sea waves sounds in water category and birdsongs and urban sounds have mean slope values around -1.0; and stream and river sounds have

mean slope values of about -0.5 to 0.0. The deviations of sea waves sounds are higher than those of wind, birdsongs, urban, and stream and river sounds. In the range of [1Hz-10Hz], wind sounds have mean slope value of about -0.5, while water, birdsongs and urban sounds have slope values spreading between about -1.5 and 0.0. The mean value of deviation of urban sounds is higher than that of birdsongs, and than water and wind sounds. In the range of [0.1Hz-1Hz], stream and river sounds have mean slope values of about 0.0; sea waves sounds and wind sounds have mean slope values of about -2.0 and -1.5; and birdsongs and urban sounds have mean slope values closer to -1.0. Urban sounds have higher mean deviation value than the other three categories of sounds. In the range of [0.005Hz-0.1Hz], wind sounds have mean slope value of -1.5; both water sounds and birdsongs have mean slope values of about -0.5 to 0.0; and urban sounds have mean slope value of about -1.0. Urban sounds have lower mean deviation value than the other three categories of sounds.

The 102 sound recordings are further plotted in three-dimensional coordinate systems with its axes presenting the slope indices in the three frequency ranges, shown in Figure 8.1.2 (e) and (f). From the figures and the results above, it can be seen that stream and river sounds have mean slope values of about 0.0 in all the three frequency ranges. Sea waves sounds have mean slope values of about -1.0 in the range of [1Hz-10Hz], about -2.0 in the range of [0.1Hz-1Hz], and about 0.0 in the range of [0.005Hz-0.1Hz]. Wind sounds have mean slope values of about -0.5 in the range of [1Hz-10Hz], and about -1.5 in the ranges of [0.1Hz-1Hz] and [0.005Hz-0.1Hz]. Birdsongs have mean slope values about or a little lower than -1.0 in the ranges of [1Hz-10Hz] and [0.1Hz-1Hz], and about -0.5 in the range of [0.005Hz-0.1Hz]. The recordings in urban category tend to be mixed with many of the recordings in the other three categories.

Comparing the results based on sharpness and on loudness, it can be seen the results are somehow similar in 1/f noise behaviours. The recordings in different sound categories in Figure 8.1.3 are located respectively at similar place to those based on loudness, shown in Figure 8.1.2, although recordings in different categories are somehow more mixed based on sharpness. That is, similarly to the results based on loudness, stream and river sounds in water category exhibit quick variations in sharpness in the three frequency ranges. Sea waves sounds in water category exhibit generally 1/f noise behaviour in sharpness in short range of 1-10s, slow variation in medium time range of 0.1-1s, and quick variation in long time ranges of 10-200s. Wind sounds exhibit quick variation in sharpness are in short time range of 0.1-1s, and slow variation in medium and long time ranges of 1-10s, and 10-200s. Birdsongs exhibit generally 1/f noise behaviour in short and medium time ranges, and quick variation in sharpness in long time range.

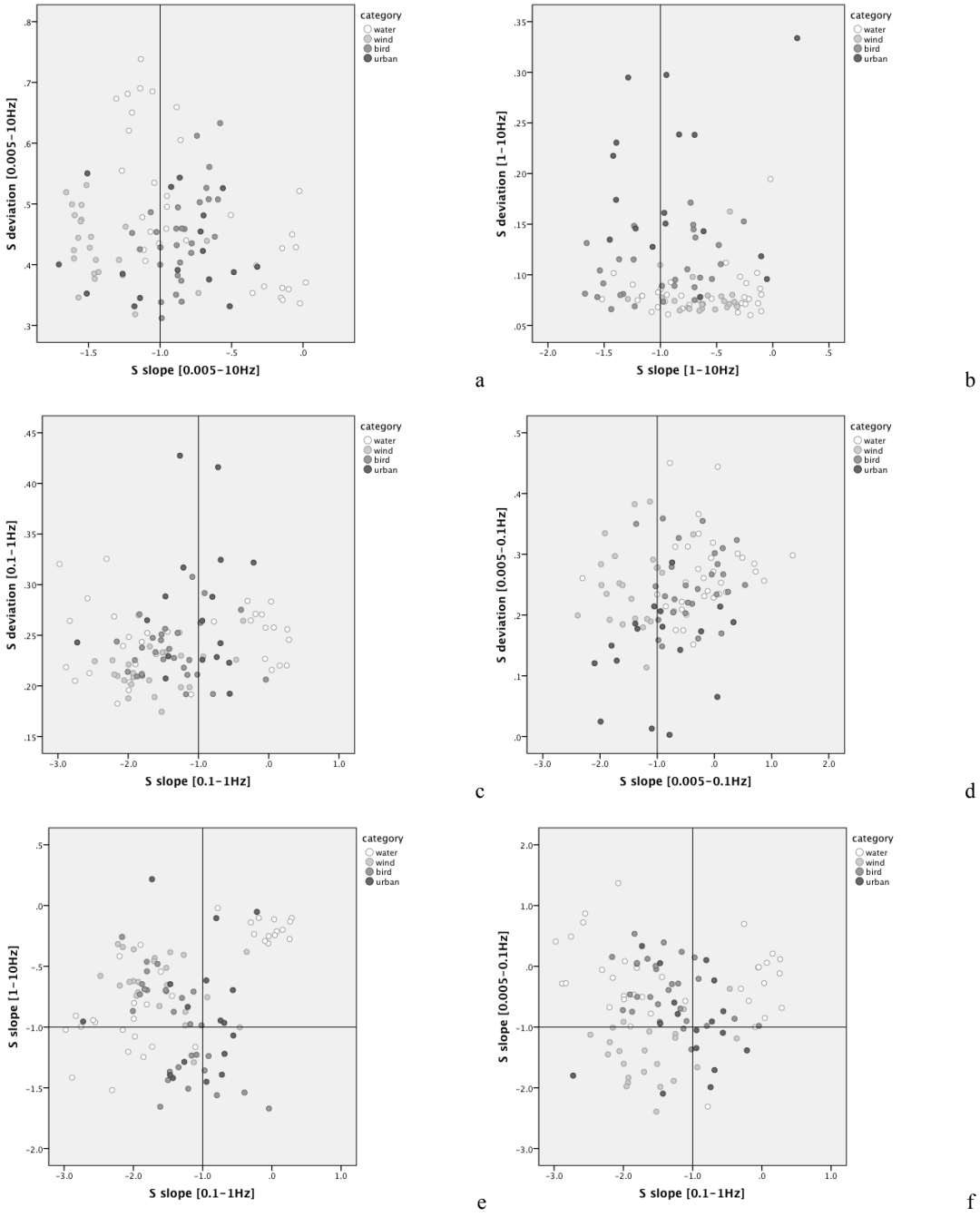


Figure 8.1.3 Characteristics of the four categories of sound in terms of 1/f noise of sharpness

8.1.3 1/f noise behaviour of tonality

1/f noise behaviours in tonality are analysed over the frequency range of 0.005-10Hz for the 102 recordings. By exploring the shapes of the spectrum density of tonality with time of the recordings, it can be seen that, different from those of loudness and sharpness, most of the shapes of spectrum density of tonality generally show a straight line, without break

points. Thus, in order to describe the shapes of spectrum density, the full frequency range of [0.005-10 Hz] is used for tonality, which responds to the time range of 0.1-200s. As a result, two indices are derived from the spectrum density, i.e., slope and deviation of spectrum density in the frequency range of [0.005-10Hz].

Table 8.1.7 Descriptives of 1/f noise indices of tonality for the four categories

		N	Mean	Std. Deviation	Minimum	Maximum		N	Mean	Std. Deviation	Minimum	Maximum
Water	Ton	34	-0.059	0.124	-0.415	0.206	Ton	34	0.364	0.040	0.297	0.451
Wind	slope	23	-0.097	0.114	-0.304	0.165	deviation	23	0.361	0.063	0.283	0.574
Bird	[0.005-	28	-0.479	0.173	-0.742	-0.085	[0.005-	28	0.408	0.083	0.290	0.668
Urban	10Hz]	17	-0.574	0.315	-1.219	-0.071	10Hz]	17	0.437	0.099	0.317	0.632

Table 8.1.8 Test of homogeneity of variances and ANOVA of 1/f noise indices of tonality for the four categories

	Test of Homogeneity of Variances		ANOVA			Test of Homogeneity of Variances		ANOVA	
	Levene Statistic	Sig.	F	Sig.		Levene Statistic	Sig.	F	Sig.
Ton slope [0.005-10Hz]	9.271	0	51.108	0	Ton deviation [0.005-10Hz]	5.418	0.002	5.972	0.001

Table 8.1.9 Multiple comparisons of 1/f noise indices of tonality for the four categories

Mean Difference (I-J)	(I) Category	Water			Wind			Bird			Urban		
Dependent Variable	(J) Category	Wind	Bird	Urban	Water	Bird	Urban	Water	Wind	Urban	Water	Wind	Bird
Ton slope [0.005-10Hz]	Dunnett T3	0.039	0.420*	0.515*	-0.039	0.381*	0.476*	-0.420*	-0.381*	0.095	-0.515*	-0.476*	-0.095
Ton deviation [0.005-10Hz]	Dunnett T3	0.003	-0.044	-0.073	-0.003	-0.047	-0.076	0.044	0.047	-0.029	0.073	0.076	0.029

The means in these 2 indices among the four categories are compared with ANOVA, to examine if the sound categories differ from each other significantly in the indices. The descriptives of the indices for the four categories are shown in Table 8.1.7, including mean, standard deviation, minimum, and maximum. The homogeneity of variances of different categories in the 1/f noise indices of tonality, as the assumption of ANOVA, is firstly tested before the ANOVA, the results of which are shown in Table 8.1.8. It shows that p values (Sig.) of both indices are less than the α level of 0.05, then the hypothesis that the variances are equal is rejected, that is, the variances are unequal. Table 8.1.8 also shows the results of ANOVA, from which it can be seen that the p values of both indices are less than the α level of 0.05, which reject the null hypothesis that all the means are equal, that is, both the indices show significant differences among the means of the categories, or between at least two categories. However, since the variances are unequal,

the results may be inaccurate, and thus further post hoc tests are checked for verifying and examining which of the specific categories differ.

The results of the post hoc tests are shown in Table 8.1.9, based on Dunnett's T3 method since the variances are unequal of the categories. It shows the pairwise multiple comparisons among the categories, in terms of difference between the means of each category with each of the other three, where asterisks indicate significantly different group means at an alpha level of 0.05. The slope of spectrum density of tonality (Ton slope [0.005-10Hz]) shows significant differences between the categories of water and wind and categories of bird and urban. As shown in Table 8.1.7, the mean slope is about -0.1 for water and wind categories, and -0.5 to -0.6 for bird and urban categories. For the deviation from the slope of spectrum density of tonality (Ton deviation [0.005-10Hz]), it does not show any significant differences among the categories. These characteristics of the four categories in terms of the indices can also be seen in Figure 8.1.4 (a), where the 102 sound recordings are plotted in the two-dimensional coordinate system with its axes presenting the slope and deviation indices of spectrum density of tonality. In other words, in the full time interval, i.e., 0.1s to 200s, bird and urban sounds exhibit relatively quick variation in tonality, while for water and wind sounds, which generally do not show any tonality (with tonality results of about 0 as discussed in Chapter 4), the mean slopes of spectrum density are equal to about 0.

8.1.4 1/f noise behaviour of pitch

In this section, variations of pitch over time are analysed based on recordings with duration of 30s, the same as that in Chapter 7. Thus, 1/f noise behaviours of the 102 recordings are calculated over the frequency range of 0.05-10Hz, which responds to the time range of 0.1-20s. By exploring the shapes of the spectrum densities of both pitch value and pitch strength, it can be seen that most of the shapes of spectrum density generally show a straight line, without break points, somehow similar to that of tonality in Section 8.1.3. Some of the spectrum densities are generally in a line, while some have break points at about 0.1Hz. Thus, to describe the shapes of spectrum density, the full frequency range of [0.05-10 Hz] is used for pitch value and pitch strength. Consequently, 4 indices of 1/f noise are derived, which are slope and deviation of spectrum density of pitch value and pitch strength in the frequency range of [0.05-10Hz].

The means among the four categories in these 4 indices are compared with ANOVA. The descriptives of the 1/f noise indices of pitch value and strength for the four categories are shown in Table 8.1.10, including mean, standard deviation, minimum and maximum.

In Table 8.1.11, the test of homogeneity of variances shows that among the indices, the p value (Sig.) of PA slope [0.05-10Hz] is greater than the α level of 0.05, which fails to reject the null hypothesis that the variances are equal, meaning that the variances are equal and the assumption of homogeneity of variance is met. For the other indices, the p values (Sig.) are less than the α level of 0.05, then the hypothesis is rejected; in other words, the variances are unequal. Also in Table 8.1.11, the results of ANOVA show that the p value (Sig.) of PA deviation [0.05-10Hz] is greater than the α level of 0.05, which fails to reject the null hypothesis that all the means are equal; for the other indices, the p values are less than 0.05, which reject the null hypothesis. That is, except for PA deviation [0.05-10Hz], the indices show significant differences among the means of the categories, or between at least two categories. However, since the variances are unequal, the results may be inaccurate, and thus further post hoc tests are checked for verifying and examining which of the specific categories differ.

The results of the post hoc tests are shown in Table 8.1.12, by either Tukey's HSD or Dunnett's T3 method according to the homogeneity of variances of the categories in the index. It shows the pairwise multiple comparisons among the categories, in terms of difference between the means of each category with each of the other three, where asterisks indicate significantly different group means at an alpha level of 0.05. The slope of spectrum density of pitch value over the range of [0.05Hz-10Hz] (PV slope [0.05-10Hz]) shows significant differences between the categories of bird and water and between bird and wind. As shown in Table 8.1.10, the mean slope of pitch value is about -0.3 for water and wind categories, and -0.5 to -0.6 for bird and urban categories. The deviation of spectrum density of pitch value over the range of [0.05Hz-10Hz] (PV slope [0.05-10Hz]) shows significant differences between the categories of wind and bird and between wind and urban. Urban category has higher mean value than bird, than water, and than wind category. For pitch strength, the slope of spectrum density (PA slope [0.05-10Hz]) shows significant differences between the categories of water and wind and categories of bird and urban. The mean slope is -0.3 for water and wind categories, and -0.6 to -0.7 for bird and urban categories. The deviation of spectrum density of pitch strength (PA deviation [0.05-10Hz]) does not show any significant differences among the categories. These characteristics of the four categories in terms of the indices can also be seen in Figure 8.1.4 (b-d), where the 102 sound recordings are plotted in the two-dimensional coordinate systems with their axes presenting the indices of slope and deviation of spectrum density.

In other words, in the time interval of 0.1s to 20s, bird and urban sounds exhibit relatively quick variations in both pitch value and pitch strength, while for water and wind sounds, the mean slopes of spectrum density are equal to about -0.3, larger than

those of bird and urban sounds, and thus exhibit even quicker variations. These results are somehow similar to those of tonality, although for tonality the characteristics of the four categories are shown in the larger time interval of 0.1s to 200s.

Table 8.1.10 Descriptives of 1/f noise indices of pitch for the four categories

		N	Mean	Std. Deviation	Minimum	Maximum		N	Mean	Std. Deviation	Minimum	Maximum
Water	PV	34	-0.261	0.280	-0.896	0.223	PV	34	0.357	0.051	0.248	0.468
Wind	slope	23	-0.307	0.218	-0.726	0.154	deviation	23	0.332	0.043	0.231	0.400
Bird	[0.05-	28	-0.601	0.290	-1.188	-0.166	[0.05-	28	0.378	0.049	0.292	0.465
Urban	10Hz]	17	-0.457	0.362	-1.073	0.106	10Hz]	17	0.399	0.075	0.295	0.526
Water	PA	34	-0.285	0.279	-0.815	0.175	PA	34	0.355	0.047	0.284	0.476
Wind	slope	23	-0.271	0.249	-0.751	0.225	deviation	23	0.343	0.038	0.280	0.422
Bird	[0.05-	28	-0.592	0.278	-1.141	-0.036	[0.05-	28	0.353	0.062	0.222	0.515
Urban	10Hz]	17	-0.652	0.277	-1.002	0.005	10Hz]	17	0.390	0.079	0.276	0.567

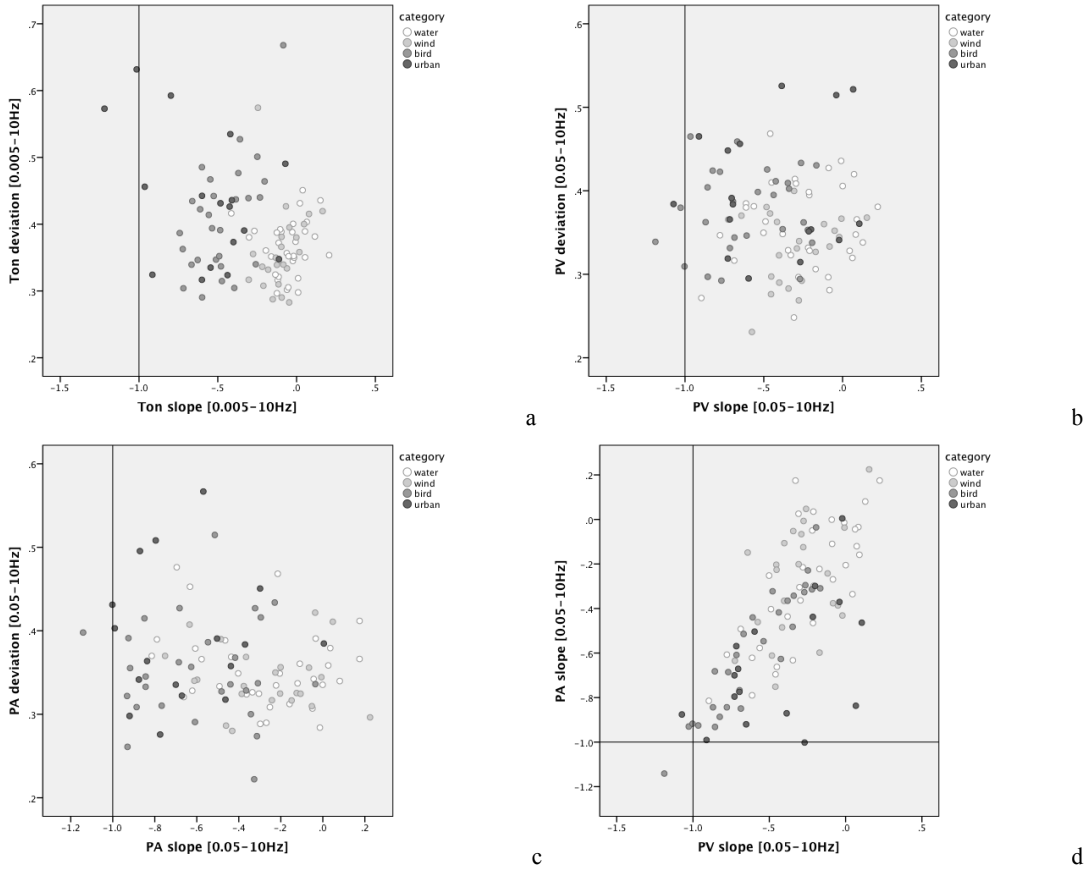


Figure 8.1.4 Characteristics of the four categories of sound in terms of 1/f noise of (a) tonality, and (b), (c) and (d) pitch

Table 8.1.11 Test of homogeneity of variances and ANOVA of 1/f noise indices of pitch for the four categories

	Test of Homogeneity of Variances		ANOVA			Test of Homogeneity of Variances		ANOVA	
	Levene Statistic	Sig.	F	Sig.		Levene Statistic	Sig.	F	Sig.
PV slope [0.05-10Hz]	3.305	0.023	8.307	0	PV deviation [0.05-10Hz]	3.478	0.019	5.986	0.001
PA slope [0.05-10Hz]	0.225	0.879	12.885	0	PA deviation [0.05-10Hz]	3.238	0.025	2.470	0.066

Table 8.1.12 Multiple comparisons of 1/f noise indices of pitch for the four categories

Mean Difference (I-J)	(I) Category	Water			Wind			Bird			Urban		
Dependent Variable	(J) Category	Wind	Bird	Urban	Water	Bird	Urban	Water	Wind	Urban	Water	Wind	Bird
PV slope [0.05-10Hz]	Dunnett T3	0.046	0.339*	0.195	-0.046	0.294*	0.150	-0.339*	-0.294*	-0.144	-0.195	-0.150	0.144
PV deviation [0.05-10Hz]	Dunnett T3	0.025	-0.020	-0.042	-0.025	-0.046*	-0.068*	0.020	0.046*	-0.022	0.042	0.068*	0.022
PA slope [0.05-10Hz]	Tukey HSD	-0.014	0.307*	0.366*	0.014	0.321*	0.380*	-0.307*	-0.321*	0.059	-0.366*	-0.380*	-0.059
PA deviation [0.05-10Hz]	Dunnett T3	0.011	0.002	-0.035	-0.011	-0.010	-0.047	-0.002	0.010	-0.037	0.035	0.047	0.037

8.1.5 Principal components of 1/f noise indices of the psychoacoustic and music parameters

The correlations between the 1/f noise indices of the psychoacoustic and music parameters are first examined based on the 102 recordings, shown in Table 8.1.13, where the indices which have high correlations with the correlation coefficients higher than 0.7 are highlighted with bold numbers. It shows that S slope [0.005-10Hz], S slope [1-10Hz] and S slope [0.1-1Hz] have high correlations (above 0.8) respectively with N slope [0.005-10Hz], N slope [1-10Hz] and N slope [0.1-1Hz]. Also, there are certain correlations (above 0.7) between N slope [0.1-1Hz] and N slope [0.005-10Hz], between N slope [0.1-1Hz] and S slope [0.005-10Hz], between S slope [0.1-1Hz] and S slope [0.005-10Hz], and between PV slope [0.05-10Hz] and PA slope [0.05-10Hz]. For deviation indices, S deviation [1-10Hz] and N deviation [1-10Hz] have high correlation. That is, slopes for 1/f noise of sharpness in the frequency ranges of [0.005-10Hz], [1-10Hz] and [0.1-1Hz] have high correlations respectively with those of loudness in the same ranges. Also, there are certain correlations between both slopes of loudness and sharpness in the range of [0.1-1Hz] and in the range of [0.005-10Hz].

Based on the 1/f noise indices of the psychoacoustic and music parameters in Section 8.1, principal component analysis (PCA) is implemented with software of SPSS Statistics 20 to reduce the dimensionality of the dataset. Based on the results of the 1/f

noise indices for the 102 recordings, the PCA is conducted on the correlation matrix of the 22 indices. Before implementation of the PCA, Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is taken, showing a result of 0.60. It generally indicates the adequacy of the sample size and the availability of the analysis.

From the 22 indices, 22 components are extracted, among which the eigenvalues of the first five components are great than one, as shown in Table 8.1.14. The first five components together account for 70.0% of the total variance. The correlations between the first five components and the indices are shown in Table 8.1.15. It can be seen that Component 1 has relatively high correlations (above 0.5) with N slope [1-10Hz], N deviation [1-10Hz], N deviation [0.1-1Hz], S slope [1-10Hz], S deviation [1-10Hz], Ton slope [0.005-10Hz], PV slope [0.05-10Hz] and PA slope [0.05-10Hz]. Component 2 has relatively high correlations with N slope [0.005-10Hz], N deviation [0.005-10Hz], N slope [0.1-1Hz], S slope [0.005-10Hz] and S slope [0.1-1Hz]. Component 3 has relatively high correlations with N slope [0.005-0.1Hz], S deviation [0.005-10Hz], S slope [0.005-0.1Hz] and S deviation [0.005-0.1Hz]. Component 4 has relatively high correlations with S deviation [0.005-10Hz], while Component 5 does not have any high correlation with any of the indices. Table 8.1.15 also shows the proportion of each index's variance that can be explained by the retained principal components. When the first five components are retained, the proportions of the indices' variance that can be explained are generally high expect for N deviation [0.005-0.1Hz], Ton deviation [0.005-10Hz] and PV deviation [0.05-10Hz], proportions of which are below 0.5. It suggests that, expect for N deviation [0.005-0.1Hz], Ton deviation [0.005-10Hz] and PV deviation [0.05-10Hz], these indices are generally well represented by the principal components.

In other words, the first component mainly represents slopes and deviations of loudness and sharpness in the range of [1-10Hz], and slopes of tonality, pitch value and pitch strength. The second component mainly represents slopes of loudness and sharpness in the ranges of [0.005-10Hz] and [0.1-1Hz]. The third component mainly represents slopes of loudness and sharpness in the range of [0.005-0.1Hz].

Thus, in order to reduce the number of indices, based on the results of both the correlation and PCA, the 1/f noise indices of sharpness, i.e. slopes and deviations of sharpness in all the frequency ranges of [0.005-10Hz], [1-10Hz] and [0.1-1Hz], and deviation indices of all the parameters can be not included in the further analysis of characteristics of sounds in Chapter 9 together with the psychoacoustic, music parameters, and 1/f noise behaviour of the parameters.

Table 8.1.13 Correlations of 1/f noise indices of psychoacoustic and music parameters

	N slope [0.005-10Hz]	N deviation [0.005-10Hz]	N slope [1-10Hz]	N deviation [1-10Hz]	N slope [0.1-1Hz]	N deviation [0.1-1Hz]	N slope [0.005-0.1Hz]	N deviation [0.005-0.1Hz]	S slope [0.005-10Hz]	S deviation [0.005-10Hz]	S slope [1-10Hz]	S deviation [1-10Hz]	S slope [0.1-1Hz]	S deviation [0.1-1Hz]	S slope [0.005-0.1Hz]	S deviation [0.005-0.1Hz]	Ton slope [0.005-10Hz]	Ton deviation [0.005-10Hz]	PV slope [0.05-10Hz]	PV deviation [0.05-10Hz]	PA slope [0.05-10Hz]	PA deviation [0.05-10Hz]
N slope [0.005-10Hz]	1.000																					
N deviation [0.005-10Hz]	-.361	1.000																				
N slope [1-10Hz]	.070	-.273	1.000																			
N deviation [1-10Hz]	.073	.155	-.273	1.000																		
N slope [0.1-1Hz]	**.776**	-.537	-.005	.165	1.000																	
N deviation [0.1-1Hz]	.003	.394	-.280	.577	.091	1.000																
N slope [0.005-0.1Hz]	.600	.168	-.283	.139	.230	.066	1.000															
N deviation [0.005-0.1Hz]	-.059	.267	.007	-.474	-.165	-.300	.032	1.000														
S slope [0.005-10Hz]	**.926**	-.380	.113	.055	**.714**	-.002	.514	-.175	1.000													
S deviation [0.005-10Hz]	-.225	.564	-.142	.016	-.495	.188	.210	.050	-.172	1.000												
S slope [1-10Hz]	.264	-.288	**.828**	-.141	.162	-.234	-.089	-.048	.325	-.125	1.000											
S deviation [1-10Hz]	.054	.035	-.185	**.914**	.140	.482	.063	-.556	.071	.029	-.064	1.000										
S slope [0.1-1Hz]	.695	-.570	.140	.055	**.900**	-.042	.121	-.207	**.727**	-.563	.239	.082	1.000									
S deviation [0.1-1Hz]	.148	.096	-.128	.398	.147	.458	.106	-.254	.197	.285	-.032	.436	.165	1.000								
S slope [0.005-0.1Hz]	.400	.086	-.207	-.065	.026	.012	.660	.132	.447	.377	-.060	-.067	-.047	.161	1.000							
S deviation [0.005-0.1Hz]	.026	-.040	.161	-.552	-.095	-.309	.086	.410	.013	.414	.109	-.533	-.140	-.135	.235	1.000						
Ton slope [0.005-10Hz]	.001	-.042	.445	-.593	-.102	-.397	-.048	.290	.088	.110	.387	-.537	-.004	-.275	.051	.407	1.000					
Ton deviation [0.005-10Hz]	.096	-.019	-.240	.341	.171	.267	.028	-.229	.059	-.041	-.209	.382	.108	.271	.024	-.241	-.285	1.000				
PV slope [0.05-10Hz]	.240	-.290	.439	-.189	.264	-.167	-.088	-.028	.304	-.150	.461	-.120	.349	.012	-.074	.126	.453	-.007	1.000			
PV deviation [0.05-10Hz]	.242	.003	-.185	.277	.279	.286	.202	-.097	.242	.042	-.155	.159	.226	.152	.138	-.026	-.198	.138	-.022	1.000		
PA slope [0.05-10Hz]	.141	-.293	.555	-.382	.108	-.351	-.105	.127	.153	-.123	.543	-.301	.184	-.137	-.066	.244	.551	-.167	**.742**	-.293	1.000	
PA deviation [0.05-10Hz]	-.009	.059	-.184	.181	.073	.141	.020	-.089	-.033	.148	-.123	.138	-.005	.317	.130	-.006	-.085	.022	-.122	.274	-.150	1.000

Table 8.1.14 Total variance explained by the components based on 1/f noise indices of psychoacoustic and music parameters

Component	Eigenvalues	% of Variance	Cumulative %
1	**4.957**	**22.532**	**22.532**
2	**4.591**	**20.866**	**43.398**
3	**2.760**	**12.545**	**55.944**
4	**1.886**	**8.571**	**64.515**
5	**1.217**	**5.534**	**70.049**
6	.995	4.521	74.570
7	.909	4.134	78.703
8	.751	3.415	82.119
9	.721	3.279	85.397
10	.638	2.898	88.295
11	.529	2.406	90.702
12	.441	2.007	92.708
13	.387	1.758	94.466
14	.307	1.396	95.862
15	.231	1.052	96.914
16	.176	.798	97.712
17	.147	.667	98.379
18	.138	.626	99.005
19	.099	.451	99.457
20	.060	.271	99.728
21	.045	.202	99.930
22	.015	.070	100.000

Table 8.1.15 Component matrix and communalities for 1/f noise indices of psychoacoustic and music parameters

	Component Matrix					Communalities
	Component 1	Component 2	Component 3	Component 4	Component 5	Extraction of 5 components
N slope [0.005-10Hz]	0.164	**0.845**	0.394	-0.067	-0.138	0.920
N deviation [0.005-10Hz]	-0.404	**-0.542**	0.309	0.275	-0.135	0.647
N slope [1-10Hz]	**0.657**	0.088	-0.320	0.441	-0.056	0.740
N deviation [1-10Hz]	**-0.765**	0.336	-0.234	0.254	-0.136	0.836
N slope [0.1-1Hz]	0.092	**0.888**	0.006	-0.242	0.159	0.881
N deviation [0.1-1Hz]	**-0.670**	0.144	-0.013	0.307	0.061	0.568
N slope [0.005-0.1Hz]	-0.145	0.348	**0.739**	-0.002	-0.360	0.818
N deviation [0.005-0.1Hz]	0.364	-0.386	0.383	-0.256	0.073	**0.499**
S slope [0.005-10Hz]	0.192	**0.850**	0.366	0.035	-0.129	0.911
S deviation [0.005-10Hz]	-0.222	-0.431	**0.539**	**0.551**	0.042	0.831
S slope [1-10Hz]	**0.594**	0.284	-0.202	0.498	-0.135	0.740
S deviation [1-10Hz]	**-0.691**	0.353	-0.310	0.330	-0.171	0.837
S slope [0.1-1Hz]	0.213	**0.869**	-0.104	-0.211	0.150	0.878
S deviation [0.1-1Hz]	-0.438	0.304	0.102	0.489	0.261	0.602
S slope [0.005-0.1Hz]	-0.052	0.167	**0.809**	0.112	-0.183	0.731
S deviation [0.005-0.1Hz]	0.481	-0.250	**0.511**	0.072	0.276	0.636
Ton slope [0.005-10Hz]	**0.734**	-0.158	0.168	0.236	0.088	0.656
Ton deviation [0.005-10Hz]	-0.408	0.258	-0.110	0.033	-0.012	**0.246**
PV slope [0.05-10Hz]	**0.565**	0.336	-0.149	0.394	0.107	0.621
PV deviation [0.05-10Hz]	-0.312	0.330	0.239	-0.057	0.463	**0.482**
PA slope [0.05-10Hz]	**0.742**	0.125	-0.130	0.357	-0.009	0.711
PA deviation [0.05-10Hz]	-0.279	0.061	0.160	0.144	0.703	0.622

8.1.6 Summary

The slopes of 1/f noise of loudness over different ranges, [1Hz-10Hz], [0.1Hz-1Hz], and [0.005Hz-0.1Hz], show the differences among water, wind and birdsongs. The sounds in urban category are mixed with many of the sounds in the other three categories. Sounds of stream and river and sounds of sea waves in water category exhibit clear differences in terms of the slopes. The slopes of stream and river sounds in water category are larger than -1 (about -0.5 to 0) for the three ranges. The slopes of sea waves sounds in water category are smaller than -1 (about -3 to -1) for the ranges of [1Hz-10Hz] and [0.1Hz-1Hz], and larger than -1 (about -0.5 to 0) for the range of [0.005Hz-0.1Hz]. The slopes of wind sounds are about -0.5, -2 and -1.5 by average for the three ranges. The slopes of birdsongs are about -1 for the ranges of [1Hz-10Hz] and [0.1Hz-1Hz], and -0.5 for the range of [0.005Hz-0.1Hz].

Comparing to the results of loudness, the results of sharpness show the similar tendencies. However, the differences among the categories are not that clear, the sounds are more mixed with the slopes of 1/f noise of sharpness over different ranges. Similar to loudness, the slopes of stream and river sounds in water category are about 0 for the three ranges. The slopes of sea waves sounds in water category are about -1, -2 and 0 by average for the three ranges. The slopes of wind sounds are about -0.5, -2 and -1.5 by average for the three ranges. The slopes of birdsongs are about -1 for the ranges of [1Hz-10Hz] and [0.1Hz-1Hz], and -0.5 for the range of [0.005Hz-0.1Hz].

The slope of spectrum density of tonality over the range of [0.005Hz-10Hz] shows significant differences between set of water and wind and set of bird and urban. The mean slopes are about 0 for water and wind, and -0.5 for bird and urban sound.

While the slope of spectrum density of pitch over the range of [0.05Hz-0.1Hz] does not show significant differences between any of the categories, slope over the range of [0.1Hz-10Hz] shows the significant differences between bird and water, and between bird and wind. Water and wind sounds have average values of both about -0.3, and birdsongs are about -0.6.

In sum, slopes of 1/f noise show that stream and river sounds in water category exhibit quick variation in loudness and sharpness in the three time intervals, i.e., 0.1s to 1s, 1s to 10s, and 10s to 200s. Sea waves sounds in water category exhibit slow variation in loudness and sharpness in short and medium time intervals of 0.1s to 1s, and 1s to 10s, and exhibit quick variation in long time intervals of 10s to 200s. Wind sounds exhibit quick variation in loudness and sharpness are in short time interval of 0.1s to 1s, and slow variation in medium and long time intervals. Birdsongs exhibit generally 1/f noise

behaviour in short and medium time intervals, and quick variation in loudness and sharpness in long time interval. For tonality, in the full time interval, i.e., 0.1s to 200s, bird and urban sounds exhibit relative quick variation in tonality, while water and wind sounds generally do not show tonality (in Chapter 4), with the mean slopes of spectrum density equal to about 0. For pitch, the variations in the full time interval of 0.1s to 20s, roughly corresponding to short and medium time intervals of loudness and sharpness, and in short time interval of 0.1s to 1s generally show similar results. Water and wind sounds have average values of both about -0.3, and birdsongs are about -0.6. In other words, water and wind sounds exhibit slightly quicker variations in pitch than birdsongs. For slopes of 1/f noise of all the four aspects, loudness, sharpness, tonality and pitch, the urban sounds vary, which are generally mixed with many of the sounds in the other three categories.

These results somehow confirm the theoretical expectations of 1/f noise in soundscape investigated by De Coensel et al. (2003), although with some divergences. De Coensel et al. (2003) found that self-organized criticality (SOC) is common in many of the activities that together generate the rural and urban soundscape, which theoretically led to the conclusion that a linear behaviour on a log-log scale of the power spectral density of loudness and pitch fluctuations could be observed in these settings.

For the two types of wind noise that contribute to outdoor soundscape: intrinsic turbulence in the air flow (pseudo-noise), and indirect sound caused by rustling grass, leafs, etc., an approximate 1/f dependence of the local wind velocity fluctuation power spectrum could be obtained from theoretical considerations, and verified by experimental data that the wind velocity fluctuations observed at a fixed location in terms of long-term variations (seconds to minutes) (De Coensel *et al.* 2003). In addition, from the relation between wind speed and wind induced noise that proved for pseudo-noise and gathered in open grassland, a single tree and several forest edges of deciduous and coniferous species, it was generally concluded that wind induced sound pressure *"is on average proportional to V^{α}, where V is the average wind speed and α is a coefficient somewhere between 1.1 and 2"* (De Coensel *et al.* 2003). Thus, based on wind speed dynamics and the relation between wind speed and wind induced noise, 1/f dependence was expected for the sound level power spectrum of pseudo-noise, while this dependence approximated $1/f^2$ more for wind induced vegetation sound level (De Coensel *et al.* 2003). In this study, wind induced vegetation sounds exhibit slopes of -2 to -1.5 in loudness in the time interval of 1s to 200s, which accords with the theoretical considerations by De Coensel et al.

For bird song, 1/f dynamics of loudness and pitch fluctuation were expected (De Coensel *et al.* 2003), since in a number of hypotheses concerning the origin of burst of bird singing (i.e. dawn chorus), the necessary ingredients were present to develop SOC,

and moreover, dawn singing directly involved sound in the creation of SOC. In this study, it is found that birdsongs have slopes of about -1 in loudness in time interval of 0.1s to 10s (but about -0.5 in the time interval of 10s to 200s), and slopes of about -0.6 in pitch in time interval of 0.1s to 20s. The results confirm that birdsongs exhibit generally 1/f noise behaviour of loudness in short to medium time interval, but relatively quick variation of pitch.

In terms of water, as the log-log linear behaviour has been observed in the flow of rivers, De Coensel et al. (2003) assumed that "the power spectrum of the sound level fluctuations of the sound observed near running or falling water exhibits linear log-log behaviour as well", however, in the relatively short time interval (0.1s to 200s) in this study, stream and river sounds have slopes of about -0.5 to 0, i.e., they exhibit quicker variation in loudness than 1/f noise.

8.2 1/f Noise Behaviour of Specific Loudness in Environmental Sounds

In this section, in addition to 1/f noise behaviours of loudness, sharpness, tonality and pitch as additional statistic index to those used in Chapters 4 and 7, 1/f noise behaviours of signal in each critical band related to the auditory system are analysed based on specific loudness. That is, it explores 1/f noise behaviour in the variation of amplitude of oscillations of a particular part of the basilar membrane (BM) in the ear, or in the activity of neural signals in the auditory nervous system.

8.2.1 Critical bands calculated and frequency range of spectrum density

Specific loudnesses of critical bands of the 102 recordings are calculated with the ArtemiS software, as described in Chapter 3. 20 bands are considered here, with the frequency borders (include start and end points) at 9.4, 90.5, 180.7, 279.2, 348.6, 442.5, 558.8,717.8, 907.6, 1113.7, 1366.6, 1746.7, 2185.2, 2725.0, 3508.8, 4518.0, 5685.4, 7388.6, 9513.9, 11570.5 and 15471.8 Hz. While a number of the recordings are low-cut filtered, the first band is not included for the calculation of 1/f noise.

As the time internal of data available of specific loudness is 0.40 seconds, the frequency range of [1-10Hz] in loudness analysis in Section 8.1.1, which responses to time internal range of [0.1-1s], is not calculated here. The shapes of spectrum density of the 102 recordings are examined over the frequency range of 0.005-1Hz consequently, which show that breaks often occur at points corresponding to about -1.0 on abscissa,

similar to that of loudness. Thus, 1/f noise behaviours of specific loudness are analysed over two ranges, i.e., [0.005-0.1Hz] and [0.1-1Hz], as well as the combined range of [0.005-1Hz], which respond to the time internal ranges of [10-200s], [1-10s], and [1-200s] respectively. In each frequency range, the slope of the spectrum density and its corresponding deviation are calculated. That is, for each of the 19 bands, 6 indices are derived from the spectrum density, i.e., slope and deviation of frequency ranges of [0.005-0.1Hz], [0.1-1Hz], and [0.005-1Hz] which as a whole. In total, 114 indices, which are slope and deviation in the three frequency ranges of the 19 bands, are calculated for the 102 recordings.

8.2.2 Correlation of the 1/f noise indices

The correlations between the 1/f noise indices of specific loudness are first examined; part of the results is shown in Table 8.2.1, Table 8.2.2 and Table 8.2.3. For correlations between the slope indices of spectrum density, it shows that generally within a certain frequency range the slopes of adjacent bands have high correlations, with correlation coefficients higher than 0.8. For example, for the full range of [0.005-1Hz] as shown in Table 8.2.1, generally the four to eight adjacent bands have high correlations (coefficient higher than 0.8). For the range of [0.1-1Hz] shown in Table 8.2.2, generally the adjacent three to seven bands have high correlations. For the range of [0.005-0.1Hz] shown in Table 8.2.3, the adjacent two bands generally have high correlations. These results suggest that generally the variations of specific loudness in adjacent bands have similar slopes of spectrum density, and the number of adjacent bands that have high correlations depends on the frequency range. For slope indices across different frequency ranges, the correlations between are generally not very high. Between the ranges of [0.005-1Hz] and [0.1-1Hz], the correlation coefficients are all below 0.5. Between the ranges of [0.1-1Hz] and [0.005-0.1Hz], the correlation coefficients are between -0.2 and 0.2. Between the ranges of [0.005-1Hz] and [0.005-0.1Hz], the correlation coefficients are between 0.3 and 0.8. The relatively high correlations exist between slope indices of a certain band, e.g. N2 slope [0.005-1Hz] and N2 slope [0.005-0.1Hz], with the correlation coefficients of around 0.75 for the 2^{nd} to 15^{th} bands and around 0.60 for the 16^{th} to 20^{th} bands.

For correlations between the deviation indices of spectrum density, within a certain frequency range, adjacent bands generally have relatively high correlations of about 0.5 to 0.8 for the ranges of [0.005-1Hz] and [0.005-0.1Hz]. For the range of [0.1-1Hz], the correlations between the deviation indices are generally below 0.6. Across different frequency ranges, the correlation coefficients are all below 0.5 between the ranges of

[0.005-1Hz] and [0.1-1Hz]. Between the ranges of [0.1-1Hz] and [0.005-0.1Hz], the correlation coefficients are between about -0.3 and 0.1. Between the ranges of [0.005-1Hz] and [0.005-0.1Hz], the indices of a certain band have relatively high correlations, with correlation coefficients of 0.6 to 0.8. The correlations between slope indices and deviation indices are generally not high within or across the three frequency ranges, with correlation coefficients of -0.3 to 0.3, except for those between the slope indices of the range of [0.1-1Hz] and the deviation indices of the full range of [0.005-1Hz], the correlation coefficients of which are between -0.6 and -0.2.

Generally, these results suggest that within a certain frequency range, the variations of specific loudness in adjacent bands show similar slopes and also similar deviations of spectrum density. Between the ranges of [0.005-1Hz] and [0.005-0.1Hz], both the slope indices of a certain band show certain correlations, as well as the deviation indices of a certain band, e.g., N2 slope [0.005-1Hz] and N2 slope [0.005-0.1Hz], and N2 deviation [0.005-1Hz] and N2 deviation [0.005-0.1Hz]. That is, the variations of specific loudness have somewhat similar tendency in spectrum density in the frequency ranges of [0.005-1Hz] and [0.005-0.1Hz].

In addition, the correlations between slope and deviation indices of specific loudness of the 19 bands and those of loudness (as in Section 8.1.1) are examined; part of the results is shown in Table 8.2.4. For the slopes, it can be seen that for the same frequency ranges, i.e., the ranges of [0.1-1Hz] and [0.005-0.1Hz], the slopes of specific loudness have certain correlations with those of loudness, with correlation coefficients of about 0.6 to 0.8. For the range of [0.005-1Hz], the slopes of specific loudness have certain correlations with those of loudness of the ranges of [0.005-10Hz] and [0.005-0.1Hz]. The correlation coefficients are 0.7 to 0.9 between the ranges of [0.005-1Hz] and [0.005-10Hz], and about 0.6 to 0.7 between the ranges of [0.005-1Hz] and [0.005-0.1Hz]. For the rest correlations among, the correlation coefficients are generally below 0.6. In terms of the deviation indices, the correlations among are generally not very high, with correlation coefficients generally lower than 0.6. The relatively high correlations (with correlation coefficients of 0.5 to 0.6) exist between the same frequency ranges, and between the slopes of specific loudness in the ranges of [0.005-1Hz] and those of loudness in the ranges of [0.005-10Hz]. Between slope and deviation indices, the correlations are generally not significant, with correlation coefficients of -0.6 to 0.3. In sum, both the slope and deviation indices of specific loudness of the 19 bands have certain correlations with those of loudness within the same frequency range, i.e., the ranges of [0.1-1Hz] and [0.005-0.1Hz]; and both the slope and deviation indices of specific loudness in the ranges of [0.005-1Hz] have certain correlations with those of loudness in the ranges of [0.005-10Hz]. The correlations between the slopes of other different ranges, between the

deviations of other different ranges, and between slopes and deviations are generally not significant.

8.2.3 1/f noise of specific loudness

In order to examine whether variations of specific loudness in frequency bands exhibit 1/f noise behaviour and whether they exhibit more likely 1/f noise behaviour than variations of loudness or less, single sample t-test is used here which compares the mean of the sample in terms of the slope indices to -1, or say, tests the null hypothesis that the population mean is equal to -1.

Table 8.2.5 shows the mean and standard deviation of the variables (or indices), as well as the Student t-statistic (t), its p-value (Sig (2-tailed)) and mean difference (the difference between the sample mean and the test value of -1). The t-statistic is the ratio of the mean difference to the standard error of the mean (estimated as the standard deviation of the sample divided by the square root of sample size). It shows that the two-tailed p-value associated with the t-test is greater the alpha level of 0.05 for the slopes of specific loudnesses of the 9^{th} to 20^{th} bands in the frequency range of [0.1-1Hz], then the null hypothesis is not rejected and it can be concluded that the mean is not significantly different from the hypothesized value of -1. For the other slope indices of specific loudness, the p-value is smaller than 0.05, so the null hypothesis is rejected and the mean is statistically different from -1. In other words, for the sample of all the four sound categories, the variations of specific loudness of the 9^{th} to 20^{th} bands in the frequency range of [0.1-1Hz] may show the 1/f noise behaviours, while the others may not.

For the slopes of loudness, it can be seen from the table that slope of loudness in the frequency range of [0.005-10Hz] (N slope [0.005-10Hz]) has a p-value greater the alpha level of 0.05, which means that the mean is not significantly different from the hypothesized value of -1. For the other slope indices of loudness, the p-value is smaller than 0.05, so the mean is statistically different from -1. That is, among the slope indices of loudness, the variation of loudness in the frequency range of [0.005-10Hz] may show the 1/f noise behaviour. Comparing the results of specific loudnesses and loudness, the t-test does not show if the variations of specific loudnesses or variations of loudness are more likely to exhibit 1/f noise behaviour. They may be generally equal in reflecting 1/f noise behaviours for the sample.

Table 8.2.1 Correlations of 1/f noise slope indices of specific loudness in the full range of [0.005-1Hz]

Pearson Correlation	N2 slope [0.005 -1Hz]	N3 slope [0.005 -1Hz]	N4 slope [0.005 -1Hz]	N5 slope [0.005 -1Hz]	N6 slope [0.005 -1Hz]	N7 slope [0.005 -1Hz]	N8 slope [0.005 -1Hz]	N9 slope [0.005 -1Hz]	N10 slope [0.005 -1Hz]	N11 slope [0.005 -1Hz]	N12 slope [0.005 -1Hz]	N13 slope [0.005 -1Hz]	N14 slope [0.005 -1Hz]	N15 slope [0.005 -1Hz]	N16 slope [0.005 -1Hz]	N17 slope [0.005 -1Hz]	N18 slope [0.005 -1Hz]	N19 slope [0.005 -1Hz]	N20 slope [0.005 -1Hz]
N2 slope [0.005-1Hz]	1																		
N3 slope [0.005-1Hz]	**.828****	1																	
N4 slope [0.005-1Hz]	**.862****	**.851****	1																
N5 slope [0.005-1Hz]	**.818****	.777**	**.913****	1															
N6 slope [0.005-1Hz]	.764**	.785**	**.864****	**.911****	1														
N7 slope [0.005-1Hz]	.778**	**.815****	**.840****	**.880****	**.950****	1													
N8 slope [0.005-1Hz]	.763**	**.809****	.796**	**.830****	**.905****	**.944****	1												
N9 slope [0.005-1Hz]	.788**	**.810****	**.801****	**.825****	**.866****	**.899****	**.958****	1											
N10 slope [0.005-1Hz]	.777**	.790**	.768**	.784**	**.813****	**.853****	**.920****	**.966****	1										
N11 slope [0.005-1Hz]	.743**	.791**	.747**	.734**	.786**	**.846****	**.894****	**.938****	**.946****	1									
N12 slope [0.005-1Hz]	.735**	.789**	.705**	.705**	.744**	**.820****	**.849****	**.900****	**.905****	**.947****	1								
N13 slope [0.005-1Hz]	.712**	.763**	.676**	.692**	.728**	**.800****	**.840****	**.890****	**.893****	**.927****	**.957****	1							
N14 slope [0.005-1Hz]	.669**	.743**	.651**	.644**	.714**	.786**	**.833****	**.869****	**.877****	**.929****	**.915****	**.929****	1						
N15 slope [0.005-1Hz]	.684**	.754**	.656**	.654**	.695**	.777**	**.812****	**.852****	**.863****	**.895****	**.905****	**.917****	**.952****	1					
N16 slope [0.005-1Hz]	.673**	.700**	.635**	.644**	.673**	.751**	.785**	**.815****	**.827****	**.857****	**.841****	**.872****	**.918****	**.955****	1				
N17 slope [0.005-1Hz]	.700**	.682**	.689**	.708**	.711**	.763**	.767**	.787**	.791**	**.818****	**.806****	**.811****	**.851****	**.889****	**.944****	1			
N18 slope [0.005-1Hz]	.682**	.656**	.701**	.719**	.701**	.735**	.739**	.751**	.762**	.776**	.751**	.752**	.796**	**.840****	**.879****	**.954****	1		
N19 slope [0.005-1Hz]	.686**	.669**	.713**	.722**	.700**	.723**	.717**	.732**	.732**	.744**	.725**	.706**	.761**	**.804****	**.844****	**.929****	**.969****	1	
N20 slope [0.005-1Hz]	.671**	.671**	.689**	.715**	.679**	.706**	.703**	.707**	.714**	.726**	.700**	.699**	.741**	.784**	**.838****	**.917****	**.947****	**.971****	1

Table 8.2.2 Correlations of 1/f noise slope indices of specific loudness in the range of [0.1-1Hz]

Pearson Correlation	N2 slope [0.1-1Hz]	**N3 slope [0.1-1Hz]**	N4 slope [0.1-1Hz]	N5 slope [0.1-1Hz]	N6 slope [0.1-1Hz]	N7 slope [0.1-1Hz]	**N8 slope [0.1-1Hz]**	N9 slope [0.1-1Hz]	N10 slope [0.1-1Hz]	N11 slope [0.1-1Hz]	N12 slope [0.1-1Hz]	**N13 slope [0.1-1Hz]**	N14 slope [0.1-1Hz]	N15 slope [0.1-1Hz]	N16 slope [0.1-1Hz]	N17 slope [0.1-1Hz]	**N18 slope [0.1-1Hz]**	N19 slope [0.1-1Hz]	N20 slope [0.1-1Hz]
N2 slope [0.1-1Hz]	1																		
N3 slope [0.1-1Hz]	.786**	1																	
N4 slope [0.1-1Hz]	.783**	**.893****	1																
N5 slope [0.1-1Hz]	.782**	**.858****	**.918****	1															
N6 slope [0.1-1Hz]	.763**	**.862****	**.866****	**.912****	1														
N7 slope [0.1-1Hz]	.769**	**.814****	**.800****	**.840****	**.924****	1													
N8 slope [0.1-1Hz]	.774**	**.805****	.782**	**.822****	**.897****	**.930****	1												
N9 slope [0.1-1Hz]	.762**	.764**	.762**	**.825****	**.867****	**.887****	**.926****	1											
N10 slope [0.1-1Hz]	.760**	.745**	.726**	.783**	**.820****	**.832****	**.887****	**.941****	1										
N11 slope [0.1-1Hz]	.792**	.780**	.734**	.786**	**.803****	**.827****	**.858****	**.900****	**.951****	1									
N12 slope [0.1-1Hz]	.696**	.755**	.749**	.785**	**.814****	**.823****	**.813****	**.855****	**.880****	**.907****	1								
N13 slope [0.1-1Hz]	.657**	.742**	.760**	.761**	.762**	.737**	.720**	.771**	.780**	**.834****	**.913****	1							
N14 slope [0.1-1Hz]	.627**	.723**	.736**	.706**	.720**	.710**	.705**	.711**	.718**	.752**	**.849****	**.922****	1						
N15 slope [0.1-1Hz]	.624**	.725**	.755**	.710**	.694**	.674**	.679**	.688**	.712**	.728**	**.826****	**.884****	**.938****	1					
N16 slope [0.1-1Hz]	.610**	.704**	.752**	.701**	.652**	.641**	.620**	.642**	.662**	.681**	.765**	**.843****	**.907****	**.953****	1				
N17 slope [0.1-1Hz]	.615**	.682**	.753**	.699**	.655**	.637**	.624**	.623**	.643**	.666**	.760**	**.814****	**.884****	**.918****	**.963****	1			
N18 slope [0.1-1Hz]	.590**	.631**	.716**	.655**	.601**	.604**	.602**	.592**	.613**	.630**	.687**	.745**	**.822****	**.868****	**.923****	**.967****	1		
N19 slope [0.1-1Hz]	.587**	.643**	.734**	.655**	.627**	.620**	.608**	.607**	.620**	.631**	.691**	.744**	**.816****	**.853****	**.906****	**.953****	**.976****	1	
N20 slope [0.1-1Hz]	.570**	.620**	.711**	.650**	.608**	.605**	.595**	.596**	.605**	.623**	.680**	.732**	**.805****	**.837****	**.889****	**.937****	**.952****	**.980****	1

Table 8.2.3 Correlations of 1/f noise slope indices of specific loudness in the range of [0.005-0.1Hz]

Pearson Correlation	N2 slope [0.005-0.1Hz]	N3 slope [0.005-0.1Hz]	N4 slope [0.005-0.1Hz]	N5 slope [0.005-0.1Hz]	N6 slope [0.005-0.1Hz]	N7 slope [0.005-0.1Hz]	N8 slope [0.005-0.1Hz]	N9 slope [0.005-0.1Hz]	N10 slope [0.005-0.1Hz]	N11 slope [0.005-0.1Hz]	N12 slope [0.005-0.1Hz]	N13 slope [0.005-0.1Hz]	N14 slope [0.005-0.1Hz]	N15 slope [0.005-0.1Hz]	N16 slope [0.005-0.1Hz]	N17 slope [0.005-0.1Hz]	N18 slope [0.005-0.1Hz]	N19 slope [0.005-0.1Hz]	N20 slope [0.005-0.1Hz]
N2 slope [0.005-0.1Hz]	1																		
N3 slope [0.005-0.1Hz]	.675**	1																	
N4 slope [0.005-0.1Hz]	.664**	**.804****	1																
N5 slope [0.005-0.1Hz]	.539**	.708**	**.820****	1															
N6 slope [0.005-0.1Hz]	.441**	.635**	.760**	.726**	1														
N7 slope [0.005-0.1Hz]	.545**	.649**	.756**	.729**	**.809****	1													
N8 slope [0.005-0.1Hz]	.525**	.676**	.696**	.670**	.641**	.784**	1												
N9 slope [0.005-0.1Hz]	.539**	.531**	.648**	.615**	.609**	.630**	**.804****	1											
N10 slope [0.005-0.1Hz]	.527**	.505**	.586**	.547**	.525**	.540**	.681**	**.853****	1										
N11 slope [0.005-0.1Hz]	.574**	.515**	.594**	.488**	.629**	.599**	.645**	.729**	.744**	1									
N12 slope [0.005-0.1Hz]	.653**	.537**	.552**	.502**	.553**	.594**	.559**	.514**	.454**	.796**	1								
N13 slope [0.005-0.1Hz]	.579**	.500**	.519**	.516**	.514**	.520**	.511**	.501**	.454**	.678**	**.863****	1							
N14 slope [0.005-0.1Hz]	.426**	.538**	.442**	.459**	.452**	.473**	.479**	.342**	.284**	.512**	.738**	.792**	1						
N15 slope [0.005-0.1Hz]	.455**	.500**	.502**	.488**	.430**	.460**	.523**	.472**	.423**	.559**	.737**	**.801****	**.887****	1					
N16 slope [0.005-0.1Hz]	.440**	.414**	.374**	.418**	.347**	.404**	.525**	.458**	.379**	.512**	.699**	.731**	**.819****	**.851****	1				
N17 slope [0.005-0.1Hz]	.543**	.386**	.395**	.403**	.362**	.402**	.496**	.466**	.367**	.559**	.728**	.729**	.682**	.703**	**.896****	1			
N18 slope [0.005-0.1Hz]	.550**	.369**	.365**	.350**	.257**	.362**	.513**	.465**	.406**	.518**	.650**	.657**	.604**	.660**	**.821****	**.934****	1		
N19 slope [0.005-0.1Hz]	.585**	.431**	.433**	.400**	.355**	.366**	.458**	.407**	.362**	.520**	.677**	.688**	.599**	.603**	.781**	**.923****	**.921****	1	
N20 slope [0.005-0.1Hz]	.536**	.434**	.397**	.398**	.344**	.377**	.492**	.417**	.366**	.514**	.681**	.692**	.654**	.671**	**.831****	**.920****	**.932****	**.964****	1

Table 8.2.4 Correlations between 1/f noise indices of specific loudness and of loudness

	N slope [0.005-10Hz]	N slope [1-10Hz]	N slope [0.1-1Hz]	N slope [0.005-0.1Hz]
N2 slope [0.005-1Hz]	.676**	-0.170	.395**	.655**
N3 slope [0.005-1Hz]	**.758****	-0.067	.482**	.612**
N4 slope [0.005-1Hz]	**.717****	-0.022	.489**	.557**
N5 slope [0.005-1Hz]	**.724****	0.035	.457**	.616**
N6 slope [0.005-1Hz]	**.744****	0.011	.515**	.611**
N7 slope [0.005-1Hz]	**.798****	-0.025	.520**	.651**
N8 slope [0.005-1Hz]	**.837****	-0.109	.569**	**.703****
N9 slope [0.005-1Hz]	**.864****	-0.155	.582**	**.741****
N10 slope [0.005-1Hz]	**.839****	-.207*	.567**	**.726****
N11 slope [0.005-1Hz]	**.859****	-0.154	.558**	**.712****
N12 slope [0.005-1Hz]	**.855****	-0.152	.559**	**.710****
N13 slope [0.005-1Hz]	**.846****	-0.174	.529**	**.728****
N14 slope [0.005-1Hz]	**.854****	-0.152	.556**	**.715****
N15 slope [0.005-1Hz]	**.853****	-0.113	.549**	.695**
N16 slope [0.005-1Hz]	**.819****	-0.091	.546**	.653**
N17 slope [0.005-1Hz]	**.788****	-0.008	.532**	.580**
N18 slope [0.005-1Hz]	**.724****	-0.039	.465**	.557**
N19 slope [0.005-1Hz]	**.710****	0.018	.477**	.520**
N20 slope [0.005-1Hz]	**.709****	0.076	.457**	.506**
N2 slope [0.1-1Hz]	.514**	.348**	.596**	0.036
N3 slope [0.1-1Hz]	.544**	.318**	.672**	-0.004
N4 slope [0.1-1Hz]	.486**	.377**	.668**	-0.098
N5 slope [0.1-1Hz]	.512**	.369**	.653**	-0.066
N6 slope [0.1-1Hz]	.541**	.349**	.672**	0.014
N7 slope [0.1-1Hz]	.570**	.370**	.672**	0.083
N8 slope [0.1-1Hz]	.600**	.305**	.693**	0.092
N9 slope [0.1-1Hz]	.615**	.311**	**.724****	0.089
N10 slope [0.1-1Hz]	.685**	.279**	**.750****	0.144
N11 slope [0.1-1Hz]	.687**	.295**	**.779****	0.145
N12 slope [0.1-1Hz]	.633**	.331**	**.824****	0.037
N13 slope [0.1-1Hz]	.553**	.258**	**.830****	-0.009
N14 slope [0.1-1Hz]	.495**	.239*	**.804****	-0.064
N15 slope [0.1-1Hz]	.477**	.270**	**.776****	-0.092
N16 slope [0.1-1Hz]	.395**	.341**	**.722****	-0.184
N17 slope [0.1-1Hz]	.361**	.392**	.692**	-.206*
N18 slope [0.1-1Hz]	.308**	.347**	.641**	-.222*
N19 slope [0.1-1Hz]	.320**	.367**	.646**	-.234*
N20 slope [0.1-1Hz]	.304**	.379**	.625**	-.262**
N2 slope [0.005-0.1Hz]	.477**	-0.185	.206*	.587**
N3 slope [0.005-0.1Hz]	.491**	-.219*	.283**	.668**
N4 slope [0.005-0.1Hz]	.545**	-0.189	.319**	.670**
N5 slope [0.005-0.1Hz]	.516**	-0.161	.235*	.693**
N6 slope [0.005-0.1Hz]	.482**	-0.190	.275**	.630**
N7 slope [0.005-0.1Hz]	.524**	-.231*	.259**	**.704****
N8 slope [0.005-0.1Hz]	.529**	-.223*	.252*	**.713****
N9 slope [0.005-0.1Hz]	.563**	-0.165	.298**	**.700****
N10 slope [0.005-0.1Hz]	.488**	-0.188	.247*	.639**
N11 slope [0.005-0.1Hz]	.523**	-0.157	.230*	.649**
N12 slope [0.005-0.1Hz]	.495**	-.228*	0.172	.674**
N13 slope [0.005-0.1Hz]	.463**	-.215*	0.104	**.705****
N14 slope [0.005-0.1Hz]	.456**	-.202*	0.131	.696**
N15 slope [0.005-0.1Hz]	.521**	-0.171	0.170	**.750****
N16 slope [0.005-0.1Hz]	.481**	-0.178	0.181	.654**
N17 slope [0.005-0.1Hz]	.487**	-0.061	0.192	.563**
N18 slope [0.005-0.1Hz]	.438**	-0.048	0.124	.551**
N19 slope [0.005-0.1Hz]	.407**	-0.074	0.145	.527**
N20 slope [0.005-0.1Hz]	.431**	-0.043	0.142	.564**

Table 8.2.5 Statistics of one-sample test (test value = -1) of 1/f noise indices of specific loudness and loudness

	Mean	Std. Deviation	t	Sig. (2-tailed)	Mean Difference
N2 slope [0.005-1Hz]	-0.755	0.459	5.394	0	0.245
N3 slope [0.005-1Hz]	-0.756	0.508	4.847	0	0.244
N4 slope [0.005-1Hz]	-0.813	0.505	3.737	0	0.187
N5 slope [0.005-1Hz]	-0.810	0.507	3.776	0	0.190
N6 slope [0.005-1Hz]	-0.806	0.533	3.674	0	0.194
N7 slope [0.005-1Hz]	-0.797	0.540	3.806	0	0.203
N8 slope [0.005-1Hz]	-0.757	0.567	4.327	0	0.243
N9 slope [0.005-1Hz]	-0.736	0.585	4.565	0	0.264
N10 slope [0.005-1Hz]	-0.749	0.609	4.158	0	0.251
N11 slope [0.005-1Hz]	-0.755	0.603	4.101	0	0.245
N12 slope [0.005-1Hz]	-0.749	0.591	4.290	0	0.251
N13 slope [0.005-1Hz]	-0.741	0.571	4.575	0	0.259
N14 slope [0.005-1Hz]	-0.686	0.580	5.460	0	0.314
N15 slope [0.005-1Hz]	-0.699	0.537	5.664	0	0.301
N16 slope [0.005-1Hz]	-0.676	0.498	6.578	0	0.324
N17 slope [0.005-1Hz]	-0.645	0.493	7.275	0	0.355
N18 slope [0.005-1Hz]	-0.646	0.465	7.701	0	0.354
N19 slope [0.005-1Hz]	-0.600	0.453	8.910	0	0.400
N20 slope [0.005-1Hz]	-0.586	0.433	9.663	0	0.414
N2 slope [0.1-1Hz]	-0.737	0.642	4.133	0	0.263
N3 slope [0.1-1Hz]	-0.771	0.712	3.246	0.002	0.229
N4 slope [0.1-1Hz]	-0.825	0.686	2.571	0.012	0.175
N5 slope [0.1-1Hz]	-0.817	0.634	2.924	0.004	0.183
N6 slope [0.1-1Hz]	-0.761	0.692	3.494	0.001	0.239
N7 slope [0.1-1Hz]	-0.800	0.721	2.799	0.006	0.200
N8 slope [0.1-1Hz]	-0.828	0.752	2.314	0.023	0.172
N9 slope [0.1-1Hz]	-0.861	0.722	1.949	**0.054**	0.139
N10 slope [0.1-1Hz]	-0.914	0.755	1.155	**0.251**	0.086
N11 slope [0.1-1Hz]	-0.942	0.754	0.780	**0.437**	0.058
N12 slope [0.1-1Hz]	-1.005	0.742	-0.073	**0.942**	-0.005
N13 slope [0.1-1Hz]	-1.082	0.735	-1.132	**0.260**	-0.082
N14 slope [0.1-1Hz]	-1.038	0.777	-0.496	**0.621**	-0.038
N15 slope [0.1-1Hz]	-1.003	0.814	-0.037	**0.971**	-0.003
N16 slope [0.1-1Hz]	-0.981	0.828	0.235	**0.815**	0.019
N17 slope [0.1-1Hz]	-0.986	0.809	0.171	**0.865**	0.014
N18 slope [0.1-1Hz]	-0.947	0.793	0.670	**0.505**	0.053
N19 slope [0.1-1Hz]	-0.894	0.779	1.379	**0.171**	0.106
N20 slope [0.1-1Hz]	-0.883	0.761	1.557	**0.123**	0.117
N2 slope [0.005-0.1Hz]	-0.778	0.695	3.230	0.002	0.222
N3 slope [0.005-0.1Hz]	-0.765	0.689	3.449	0.001	0.235
N4 slope [0.005-0.1Hz]	-0.724	0.730	3.817	0	0.276
N5 slope [0.005-0.1Hz]	-0.712	0.753	3.859	0	0.288
N6 slope [0.005-0.1Hz]	-0.677	0.838	3.898	0	0.323
N7 slope [0.005-0.1Hz]	-0.700	0.727	4.173	0	0.300
N8 slope [0.005-0.1Hz]	-0.664	0.726	4.675	0	0.336
N9 slope [0.005-0.1Hz]	-0.637	0.730	5.013	0	0.363
N10 slope [0.005-0.1Hz]	-0.645	0.841	4.257	0	0.355
N11 slope [0.005-0.1Hz]	-0.697	0.773	3.957	0	0.303
N12 slope [0.005-0.1Hz]	-0.673	0.779	4.243	0	0.327
N13 slope [0.005-0.1Hz]	-0.590	0.832	4.976	0	0.410
N14 slope [0.005-0.1Hz]	-0.552	0.967	4.683	0	0.448
N15 slope [0.005-0.1Hz]	-0.603	0.837	4.787	0	0.397
N16 slope [0.005-0.1Hz]	-0.567	0.796	5.490	0	0.433
N17 slope [0.005-0.1Hz]	-0.486	0.869	5.970	0	0.514
N18 slope [0.005-0.1Hz]	-0.541	0.854	5.421	0	0.459
N19 slope [0.005-0.1Hz]	-0.480	0.884	5.937	0	0.520
N20 slope [0.005-0.1Hz]	-0.504	0.816	6.136	0	0.496
N slope [0.005-10Hz]	-1.055	0.517	-1.071	**0.287**	-0.055
N slope [1-10Hz]	-0.867	0.540	2.488	0.014	0.133
N slope [0.1-1Hz]	-1.458	0.956	-4.833	0	-0.458
N slope [0.005-0.1Hz]	-0.700	0.716	4.228	0	0.300

8.2.4 Characteristics of different categories of sounds in terms of 1/f noise behaviour of the specific loudness

As discussed above in Section 8.2.2, there are relatively high correlations within a certain frequency range between slope indices of specific loudness of adjacent bands and between deviation indices of specific loudness of adjacent bands, and relatively high correlations between the slope indices of a certain band in the frequency range of [0.005-1Hz] and range of [0.005-0.1Hz], as well as the deviation indices of a certain band. Thus, to analyse the characteristics of different categories of sound in terms of variation of specific loudness, only a number of indices are used among all the slope and deviation indices as discussed above. Here, the slopes of 3rd, 8th, 13th, and 18th bands in the frequency ranges of [0.1-1Hz] and [0.005-0.1Hz] are used.

With these 8 indices, the 102 sound recordings are plotted in the two-dimensional coordinate systems with their axes presenting respectively the indices of slope in the different ranges, shown in Figure 8.2.1. It can be seen from the plots, similarly to that of loudness discussed in Section 8.1.1, that the recordings in water sound category gather in two groups in the plots. Detailed data show that these two groups respectively are stream and river sound recordings and sea waves sound recordings. Stream and river sounds have mean slope values of about 0.0 of all the four bands in both the frequency ranges. Sea waves sounds have mean slope values of about -2.0 in the range of [0.1-1Hz] and about 0.0 in the range of [0.005-0.1Hz] of all the four bands. Wind sounds have mean slope values of about -0.5 for the 3rd and 18th bands, and about -1.0 for the 8th and 13th bands in the range of [0.1-1Hz], and have mean slope values of about -1.5 for the four bands in the range of [0.005-0.1Hz]. Birdsongs have mean slope values of about -0.5 for the 3rd and 8th bands, about -1.0 for the 13th band, and about -1.5 for the 18th band in the range of [0.1-1Hz], and have mean slope values of about -0.5 for the four bands in the range of [0.005-0.1Hz]. While the recordings in water, wind and bird categories are generally apart from each other in the plots, the recordings in urban category are more dispersive, mixed with many of the recordings in the other three categories. From the results, it can be seen that for the range of [0.005-0.1Hz], the slope values of the different bands do not differ much for all the four types of sound. For the range of [0.1-1Hz], for water and urban sounds, the slope values of the different bands do not differ much; for wind sounds and birdsongs, the slope values vary for the different bands.

Comparing these results of specific loudness in frequency bands to those of loudness in Section 8.1.1, it can be seen that for all the four types of sound, in the range of [0.005-0.1Hz], the characteristics of 1/f noise behaviour in specific loudness and those in loudness are similar. In the range of [0.1-1Hz], for water sounds, the characteristics in

specific loudness and those in loudness are also similar. For wind sounds, the mean slope values are about -0.5 for the 3rd and 18th bands and -1.0 for the 8th and 13th bands in specific loudness, and about -2.0 in loudness in the range of [0.1-1Hz]. For birdsongs, the mean slope values are about -0.5 for the 3rd and 8th bands, -1.0 for the 13th band, and -1.5 for the 18th band in specific loudness in the range of [0.1-1Hz], while the mean slope value is about -1.0 in loudness. For both specific loudness and loudness, urban sounds have relatively more wide ranges of slope value compared to the other three types of sound.

In sum, generally, for all the four categories of sound, the characteristics of 1/f noise behaviour do not differ much with different frequency bands, and also do not differ much from those of loudness, although there are some differences, e.g., in the range of [0.1-1Hz], for both wind sounds and birdsongs, the slope values vary with different bands and differ from that of loudness. That is, variations of specific loudness of different frequency bands or that of loudness do not to a large degree effect in showing 1/f noise behaviour in loudness aspect.

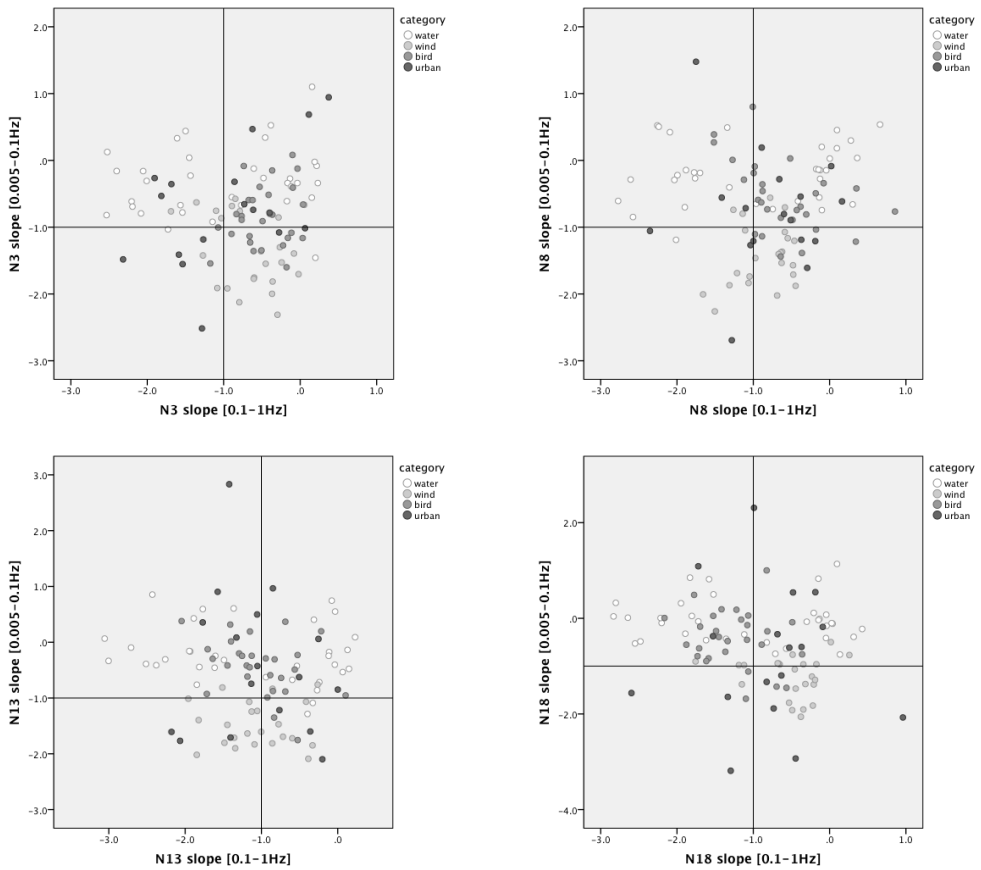


Figure 8.2.1 Characteristics of the four types of sound in terms of 1/f noise slopes indices of specific loudness in the ranges of [0.1-1Hz] and [0.005-0.1Hz]

8.3 Conclusions

The characteristics of the different categories of sound are explored in terms of the 1/f noise behaviour of the results of the parameters in Chapters 4 and 7. These parameters are loudness, sharpness, tonality and pitch, as well as specific loudnesses that simulate the variation in each critical band. The 1/f noise behaviour can be seen as a statistic index in addition to those commonly used.

Slopes of 1/f noise show that stream and river sounds in water category exhibit quick variation in loudness and sharpness in the three time intervals, i.e., 0.1s to 1s, 1s to 10s, and 10s to 200s. Sea waves sounds in water category exhibit slow variation in loudness and sharpness in short and medium time intervals of 0.1s to 1s, and 1s to 10s, and exhibit quick variation in long time interval of 10s to 200s. Wind sounds exhibit quick variations in loudness and sharpness are in short time intervals of 0.1s to 1s, and slow variations in medium and long time intervals. Birdsongs exhibit generally 1/f noise behaviour in short and medium time intervals, and quick variations in loudness and sharpness in long time intervals. For tonality, in the full time interval, i.e., 0.1s to 200s, bird and urban sounds exhibit relative quick variation in tonality, while water and wind sounds generally do not show tonality (also see Chapter 4), with the mean slopes spectrum density equal to about 0. For pitch, the variations in the time interval of 0.1s to 20s, roughly corresponds to short and medium time intervals of loudness and sharpness, and in short time interval of 0.1s to 1s generally show similar results. Water and wind sounds have average values of both about -0.3, and birdsongs are about -0.6. In other words, water and wind sounds exhibit slightly quicker variations in pitch than birdsongs. For slopes of 1/f noise of all the four aspects, loudness, sharpness, tonality and pitch, the urban sounds vary, generally mixed with many of the sounds in the other three categories. In terms of specific loudness in frequency bands, generally, it shows similar characteristics of 1/f noise behaviour for the four sound categories to those of loudness.

Chapter 9
Characteristics of natural and urban sounds in soundscapes in terms of psychoacoustic, music parameters, and 1/f noise behaviour of the parameters

In this chapter, all the parameters that have been discussed in Chapters 4, 7 and 8, which include loudness, sharpness, tonality, roughness, fluctuation strength, pitch value, pitch strength, event interval, event density, attack slope (or spectral flux) of event, periodicity, and 1/f noise behaviour in terms of these parameters, are considered together to analyse the characteristics of environmental sounds in soundscapes. In Section 9.1, principal components of all the indices in this thesis together are analysed, including mean, standard deviation, maximum and minimum, percentiles, and 1/f noise of the psychoacoustic and music parameters. In Section 9.2, characteristics of environmental sounds and differences between the different sound categories are analysed based on the results of the PCA, and are summarised from the chapters in this thesis. Finally, in Sections 9.3 and 9.4, the categories of recordings are automatically identified with all the parameters respectively using discriminant function analysis (DFA) and artificial neural network (ANN).

9.1 Principal Components of Psychoacoustic, Music Parameters, and 1/f Noise Behaviour of the Parameters

Principal components of all the indices used in last chapters, i.e., Chapters 4, 7 and 8, are explored. The principal component analysis (PCA) is implemented based on the results of the 32 indices, including the key psychoacoustic, music and 1/f noise indices, of the 102 recordings. Before the PCA, the correlations between the 1/f noise indices and the psychoacoustic and music indices are first examined, which shows that the correlations are generally not high, with the correlation coefficients lower than 0.6. Also, Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is taken, showing a result of 0.79, which indicates the adequacy of the sample size and the availability of the analysis.

From the 32 indices, 32 components are extracted, among which the eigenvalues of the first seven components are greater than one, as shown in Table 9.1.1. The first seven

components together account for 79.7% of the total variance. The correlations between the first seven components and the indices are shown in Table 9.1.2. It can be seen that Component 1 has relatively high correlations (above 0.6) with L STDEV, S STDEV, Fls AVE, Fls STDEV, PV1, PA1, PN, PV AVE, PA AVE, ED(E), AS AVE, BS and Ton slope [0.005-10Hz]. Component 2 has relatively high correlations with L AVE, N AVE, N STDEV and SF AVE. Component 3 has relatively high correlations with S AVE, R AVE, N slope [0.005-10Hz] and N slope [0.1-1Hz]. Component 4 has relatively high correlation with S slope [0.005-0.1Hz], while Components 5, 6 and 7 do not have any particularly high correlation with any of the indices. Table 9.1.2 also shows the proportion of each index's variance that can be explained by the retained principal components. When the first seven components are retained, the proportions of all the indices' variance that can be explained are generally high (all above 0.5), which means that generally all these indices are well represented by the principal components. When the first four components are retained, R STDEV, IOI (SF) AVE and N slope [0.1-1Hz], which have relatively low proportion values of below 0.5, are not well represented. Generally, the first four components may be necessary to represent the original indices, which together account for 66.2% of the total variance.

Table 9.1.1 Total variance explained by the components based on psychoacoustic, music, and 1/f noise indices

Component	Eigenvalues	% of Variance	Cumulative %
1	**10.533**	**32.916**	**32.916**
2	**4.630**	**14.468**	**47.384**
3	**3.484**	**10.888**	**58.272**
4	**2.522**	**7.883**	**66.154**
5	**1.928**	**6.026**	**72.180**
6	**1.240**	**3.877**	**76.057**
7	**1.152**	**3.599**	**79.656**
8	.982	3.069	82.725
9	.861	2.691	85.416
10	.543	1.696	87.112
11	.489	1.527	88.638
12	.462	1.445	90.084
13	.454	1.420	91.504
14	.397	1.240	92.744
15	.329	1.028	93.772
16	.299	.936	94.708
17	.247	.771	95.478
18	.241	.752	96.230
19	.202	.631	96.861
20	.168	.525	97.386
21	.161	.503	97.889
22	.127	.397	98.286

Generally, the first component mainly represents fluctuation and rhythm properties, including L STDEV, S STDEV, Fls AVE, Fls STDEV, ED(E), AS AVE and BS, and

pitch properties, including PV1, PA1, PN, PV AVE and PA AVE. The second component mainly represents loudness properties, including L AVE, N AVE and N STDEV. The third component mainly represents timbre properties, including S AVE and R AVE, and 1/f noise properties, including N slope [0.005-10Hz] and N slope [0.1-1Hz]. The fourth component mainly represents 1/f noise properties in the frequency range of [0.005-0.1Hz], including S slope [0.005-0.1Hz]. Of the indices of psychoacoustic and music parameters and 1/f noise indices of the parameters used in this study for the measurement of environmental sounds in soundscapes, the main dimensions are identified based on the sample. The first may be rhythm (includes fluctuation) and pitch; the second may be loudness; the third may be timbre and 1/f noise.

Table 9.1.2 Component matrix and communalities for psychoacoustic, music, and 1/f noise indices

	Component Matrix							Communalities	
	Component 1	Component 2	Component 3	Component 4	Component 5	Component 6	Component 7	Extraction of 4 components	Extraction of 7 components
L AVE	-0.339	**0.631**	0.586	-0.160	0.166	0.010	-0.117	0.882	0.923
L STDEV	**0.722**	0.287	-0.396	0.281	0.110	0.068	0.121	0.839	0.870
N AVE	-0.405	**0.647**	0.574	-0.034	0.098	0.109	-0.167	0.913	0.962
N STDEV	0.302	**0.862**	-0.026	0.170	0.107	0.021	0.061	0.865	0.880
S AVE	0.396	-0.064	**0.610**	0.289	0.341	0.209	-0.209	0.616	0.820
S STDEV	**0.844**	0.074	0.090	0.137	0.252	0.186	0.084	0.746	0.851
Ton AVE	0.381	0.446	0.145	-0.487	-0.428	0.221	0.084	0.601	0.840
Ton STDEV	0.511	0.407	0.094	-0.457	-0.445	0.084	0.188	0.644	0.885
R AVE	-0.489	0.436	**0.656**	0.216	0.137	-0.131	-0.033	0.906	0.943
R STDEV	0.422	0.431	0.043	0.273	0.156	-0.255	0.501	**0.441**	0.781
Fls AVE	**0.785**	-0.001	0.296	0.034	0.064	-0.022	0.357	0.705	0.837
Fls STDEV	**0.776**	0.041	0.183	0.065	0.077	-0.021	0.301	0.642	0.739
PV1	**0.732**	-0.256	0.086	0.088	0.252	0.022	-0.278	0.616	0.758
PA1	**0.827**	-0.038	0.123	-0.132	-0.101	0.019	-0.182	0.718	0.762
PN	**-0.660**	0.238	0.084	0.309	-0.197	0.249	0.200	0.595	0.736
PV AVE	**0.811**	-0.304	0.091	0.106	0.296	0.060	-0.168	0.770	0.889
PA AVE	**0.777**	-0.176	0.153	-0.167	0.091	0.024	0.027	0.687	0.696
IOI(E) AVE	-0.458	0.452	-0.200	0.213	0.119	0.232	0.045	0.500	0.570
ED(E)	**0.616**	-0.469	0.193	-0.230	0.193	-0.238	0.081	0.689	0.789
AS AVE	**0.852**	-0.192	0.106	-0.029	0.114	-0.180	0.111	0.774	0.832
IOI (SF) AVE	0.252	0.360	-0.241	0.350	-0.220	-0.534	-0.218	**0.374**	0.755
ED(SF)	-0.520	-0.342	0.300	-0.491	0.177	0.235	0.063	0.719	0.809
SF AVE	0.235	**0.788**	0.305	-0.012	0.181	-0.211	-0.176	0.769	0.878
BS	**0.664**	0.171	0.096	-0.299	-0.042	0.355	0.105	0.569	0.708
N slope [0.005-10Hz]	0.281	-0.388	**0.641**	0.277	-0.432	-0.005	-0.074	0.718	0.910
N slope [1-10Hz]	-0.541	-0.363	0.359	-0.087	0.399	0.026	0.134	0.561	0.739
N slope [0.1-1Hz]	0.145	-0.238	**0.633**	-0.117	-0.514	-0.235	-0.142	**0.492**	0.832
N slope [0.005-0.1Hz]	0.399	-0.154	0.162	0.599	-0.409	0.196	-0.025	0.568	0.775
S slope [0.005-0.1Hz]	0.286	-0.157	0.093	**0.643**	-0.222	0.347	-0.031	0.528	0.698
Ton slope [0.005-10Hz]	**-0.643**	-0.294	0.034	0.313	0.065	0.095	0.140	0.599	0.632
PV slope [0.05-10Hz]	-0.533	-0.219	0.435	0.077	-0.161	-0.231	0.356	0.527	0.733
PA slope [0.05-10Hz]	-0.581	-0.331	0.353	0.165	0.096	-0.093	0.208	0.599	0.660

9.2 Characteristics of Natural and Urban Sounds in Soundscapes in Terms of Principal Components of Psychoacoustic, Music Parameters, and 1/f Noise Behaviour of the Parameters

To summarise the results above in Chapters 4, 7 and 8, the different characteristics of the different types of environmental sound can be drawn, with a number of key psychoacoustic, music, and 1/f noise indices.

Generally, water sounds have a wide range of loudness, low pitch value and strength, low fluctuation strength, low event density (high event interval), low attack slope of event, and do not show periodicity. In water category, stream and river sounds exhibit quicker variations than 1/f noise in loudness and sharpness in the full time interval of 0.1s to 200s, while sea waves sounds exhibit relatively slow variations in loudness and sharpness in short (0.1s to 1s) and medium (1s to 10s) time intervals, and relatively quick variations in long time interval (10s to 200s).

Wind sounds have a wide range of loudness, low sharpness, low pitch value and strength, low fluctuation strength, low event density (high event interval), low attack slope of event, and do not show periodicity. Moreover, wind sounds exhibit quicker variations than 1/f noise in loudness and sharpness are in short time interval, and slower variations in medium and long time intervals.

Birdsongs have low loudness, high sharpness, high pitch value and strength, high fluctuation strength, high event density (low event interval), high attack slope of event, and may show periodicity. Birdsongs exhibit generally 1/f noise behaviour in short and medium time intervals, and relatively quick variations in loudness and sharpness in long time interval.

Urban sounds have high loudness, low pitch values and a relatively wide range of pitch strengths, a relatively wide range of fluctuation strength, event density, and attack slope of event, and may show periodicity. The 1/f noise behaviours of urban sounds also vary in a relatively wide range. The various characteristics of urban sounds may be explained by the complicated sound types in the urban sound category.

Among the sound categories, water and wind sounds have similar characteristics in terms of psychoacoustic and music parameters. For instance, both of them have a wide range of loudness, low pitch value and strength, and low fluctuation strength. However, 1/f noise indices show the differences between them. In the long time interval of 10s to 200s, water sounds, both stream and river sounds and sea waves sounds, exhibit relatively quick variations compared to 1/f noise, while wind sounds exhibit relatively slow variations.

In terms of the differences between natural and urban sounds, generally speaking, urban sounds are with high fluctuation strength and loudness, while natural sounds are either with low fluctuation strength and varied loudness, or with high fluctuation strength and low loudness.

9.3 Automatic Identification of Sound Categories with Discriminant Functions Analysis

Discriminant function analyses (DFAs) are used here to automatically identify sound categories of the 102 recordings, based on the psychoacoustic and music parameters and 1/f noise behaviour of the parameters. The discriminant functions obtained, i.e. linear combinations of the variables (or indices), can be used to predict group membership of recordings with known variables and unknown group membership. Also, discriminant functions give insight into the relationship between group membership and the variables used.

9.3.1 Discriminant function analysis based on the psychoacoustic, music, and 1/f noise indices

Based on the 32 indices as discussed in the last section, which include the key psychoacoustic, music and 1/f noise indices, DFA is implemented. All the indices are considered together, i.e., all the indices are included in the discriminant functions. As the 102 recordings are labelled in four groups, i.e. water, wind, bird and urban, three canonical linear discriminant functions are developed, one less than the number of levels in the group variable. Each function acts as a projection of the data onto a dimension that best separates or discriminates between the groups. Table 9.3.1 shows the eigenvalue, which indicates the function's discriminating ability, the proportion and cumulative proportion of discriminating ability, and the canonical correlation of predictor variables and the groupings of each given function. The first function accounts for 80.4% of the discriminating ability of the discriminating variables, and the second function accounts for 10.1%. The canonical correlations of all the three functions are high, larger than 0.8. It indicates good correlations between the functions and groupings, that is, the functions are effective for the discriminating. Table 9.3.1 also shows the tests of functions with the null hypothesis that a function, and all functions that follow, have no discriminating ability, in terms of Wilks' Lambda and Chi-square statistic. The p-values (Sig.) associated with the

Chi-square statistic are smaller than the alpha level of 0.05 for all the three tests, which rejects the null hypothesis. That is, all the three functions have discriminating ability.

Table 9.3.1 Eigenvalues and Wilks' Lambda of discriminant functions based on psychoacoustic, music, and 1/f noise indices

Eigenvalues					Wilks' Lambda			
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation	Test of Function(s)	Wilks' Lambda	Chi-square	Sig.
1	24.359	80.4	80.4	.980	1 through 3	.003	497.259	.000
2	3.058	10.1	90.5	.868	2 through 3	.063	228.909	.000
3	2.885	9.5	100.0	.862	3	.257	112.647	.000

Table 9.3.2 Structure matrix of discriminant functions based on psychoacoustic, music, and 1/f noise indices

	Function 1	Function 2	Function 3
PV AVE	.513*	-.167	.072
PV1	.309*	-.069	.044
S STDEV	.292*	-.059	-.055
PA1	.284*	.027	-.221
PA AVE	.249*	.029	-.118
AS AVE	.224*	.013	-.177
PN	-.196*	-.173	.095
ED(E)	.160*	.066	-.075
L STDEV	.111*	-.040	-.102
R AVE	-.106*	.022	-.082
PV slope [0.05-10Hz]	-.095*	-.069	.082
S slope [0.005-0.1Hz]	.045	-.437*	-.141
N slope [0.005-0.1Hz]	.063	-.428*	-.224
N slope [0.005-10Hz]	.072	-.363*	-.240
L AVE	-.083	.217*	-.126
S AVE	.125	-.137*	-.083
N AVE	-.094	.136*	-.113
Ton slope [0.005-10Hz]	-.169	-.288	.463*
Ton STDEV	.061	.226	-.369*
Ton AVE	.037	.199	-.349*
PA slope [0.05-10Hz]	-.086	-.112	.246*
SF AVE	.014	.205	-.233*
Fls AVE	.200	.052	-.227*
Fls STDEV	.180	.065	-.214*
BS	.129	.094	-.206*
N slope [1-10Hz]	-.071	.056	.201*
N slope [0.1-1Hz]	.036	-.069	-.188*
IOI(E) AVE	-.096	.054	.185*
N STDEV	.012	.114	-.164*
R STDEV	.042	.042	-.163*
ED(SF)	-.064	.119	.154*
IOI(SF) AVE	.019	-.080	-.090*

Structure matrix of the discriminant functions, shown in Table 9.3.2, displays the correlations between the discriminating variables and the dimensions created with the discriminant functions, in which the variables are ordered by absolute value of correlation

within function and * indicates largest absolute correlation between each variable and any discriminant function. The first function has relatively high correlations (larger than 0.3) with pitch values, including PV AVE and PV1. The second function has relatively high correlations with 1/f noise, including S slope [0.005-0.1Hz], N slope [0.005-0.1Hz] and N slope [0.005-10Hz]. The third function has relatively high correlations with tonality, Ton slope [0.005-10Hz], Ton STDEV and Ton AVE. Generally, the first function mainly represents pitch value; the second function mainly represents 1/f noise; and third mainly represents tonality.

Table 9.3.3 Unstandardized discriminant function coefficients based on psychoacoustic, music, and 1/f noise indices, and based on the indices with fountain sounds labelled as water sounds

	Unstandardized function coefficients			Unstandardized function coefficients (with fountain sounds labelled as water sounds)		
	Function 1	Function 2	Function 3	Function 1	Function 2	Function 3
L AVE	-.005	.008	-.017	.006	.046	-.022
L STDEV	-.391	.095	-.064	-.348	.235	-.056
N AVE	.026	.093	.083	.035	.070	.107
N STDEV	-.002	-.097	.103	-.015	-.140	.063
S AVE	.298	.032	-1.190	-.059	-.695	-.762
S STDEV	5.683	-1.599	2.593	5.345	-3.239	1.990
Ton AVE	-3.983	-.403	-10.873	-4.036	2.920	-10.029
Ton STDEV	-3.243	19.193	4.412	-.206	22.061	9.563
R AVE	-.554	-.668	-.285	-.677	-.746	-.440
R STDEV	.560	-.152	-.277	.660	.232	-.399
Fls AVE	3.805	3.215	4.101	7.831	12.813	1.675
Fls STDEV	5.319	31.798	-15.776	4.345	25.046	.026
PV1	.000	.000	.000	.000	.000	.000
PA1	1.147	-.261	.333	.936	-.970	.376
PN	-2.035	-.640	.240	-2.058	-.503	-.043
PV AVE	.003	.000	.002	.004	.000	.002
PA AVE	.465	-.533	-.886	.533	.036	-1.088
IOI(E) AVE	.092	.197	.117	.073	.050	.207
ED(E)	.048	.384	-.150	.056	.352	.020
AS AVE	-.342	-.143	.011	-.348	-.111	-.046
IOI (SF) AVE	.262	-.320	.158	.155	-.613	.100
ED(SF)	.044	-.120	.033	.027	-.151	-.007
SF AVE	-.014	.004	-.023	-.009	.024	-.024
BS	-.049	-.202	.283	-.006	-.101	.134
N slope [0.005-10Hz]	-1.288	-4.186	-3.085	-1.983	-4.184	-3.996
N slope [1-10Hz]	-.479	1.251	.653	-.310	1.262	.979
N slope [0.1-1Hz]	.656	1.086	.950	1.000	1.505	1.035
N slope [0.005-0.1Hz]	-.108	-.028	.004	-.009	.274	-.093
S slope [0.005-0.1Hz]	-.014	-.383	-.079	-.010	-.251	-.237
Ton slope [0.005-10Hz]	-1.024	-1.029	3.922	-.901	-1.487	3.019
PV slope [0.05-10Hz]	.145	1.165	-.129	-.199	-.117	.672
PA slope [0.05-10Hz]	-.332	-.405	1.452	-.122	-.089	.968
(Constant)	-2.723	-3.271	2.406	-2.857	-3.374	.900

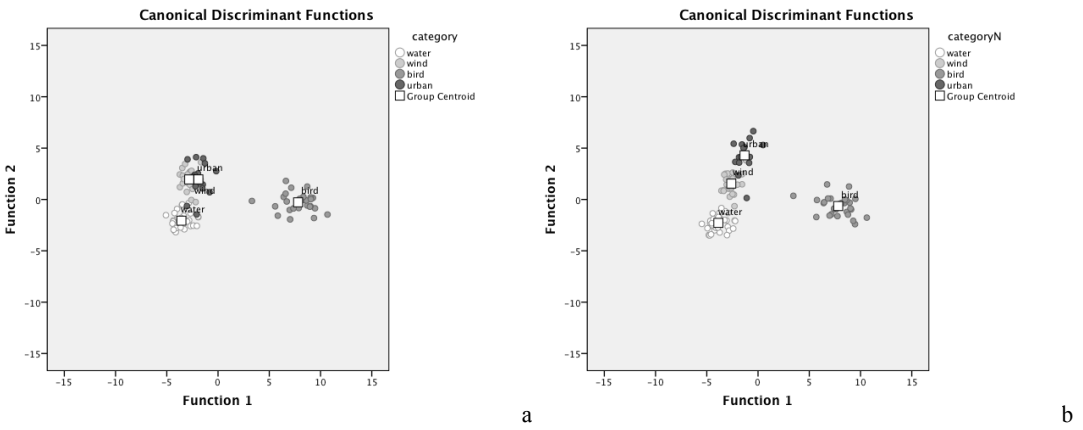
Table 9.3.4 Group centroids of discriminant functions based on psychoacoustic, music, and 1/f noise indices, and based on the indices with fountain sounds labelled as water sounds

	Group centroids of the categories			Group centroids of the categories (with fountain sounds labelled as water sounds)		
Category	Function 1	Function 2	Function 3	Function 1	Function 2	Function 3
Water	-3.574	-2.066	-.072	-3.870	-2.281	-.742
Wind	-2.811	1.950	2.236	-2.608	1.545	2.676
Bird	7.812	-.280	.154	7.824	-.640	.044
Urban	-1.916	1.955	-3.134	-1.318	4.300	-2.407

Table 9.3.5 Classification results by discriminant functions based on psychoacoustic, music and 1/f noise indices, and results based on the indices with fountain sounds labelled as water sounds

		Category	Predicted Group Membership					Predicted Group Membership (with fountain sounds labelled as water sounds)				
			Water	Wind	Bird	Urban	Total	Water	Wind	Bird	Urban	Total
Original	Count	Water	34	0	0	0	34	36	0	0	0	36
		Wind	1	22	0	0	23	1	22	0	0	23
		Bird	0	0	28	0	28	0	0	28	0	28
		Urban	1	0	0	16	17	1	0	0	14	15
	%	Water	100.0	.0	.0	.0	100.0	100.0	.0	.0	.0	100.0
		Wind	4.3	95.7	.0	.0	100.0	4.3	95.7	.0	.0	100.0
		Bird	.0	.0	100.0	.0	100.0	.0	.0	100.0	.0	100.0
		Urban	5.9	.0	.0	94.1	100.0	6.7	.0	.0	93.3	100.0
Cross-validated	Count	Water	32	2	0	0	34	36	0	0	0	36
		Wind	2	20	0	1	23	2	20	0	1	23
		Bird	0	0	27	1	28	0	0	27	1	28
		Urban	4	4	0	9	17	1	5	0	9	15
	%	Water	94.1	5.9	.0	.0	100.0	100.0	.0	.0	.0	100.0
		Wind	8.7	87.0	.0	4.3	100.0	8.7	87.0	.0	4.3	100.0
		Bird	.0	.0	96.4	3.6	100.0	.0	.0	96.4	3.6	100.0
		Urban	23.5	23.5	.0	52.9	100.0	6.7	33.3	.0	60.0	100.0

The unstandardized canonical discriminant function coefficients for each variable of each function are shown in Table 9.3.3, which are used for the classification/prediction, together with the group centroids, i.e. the means of the discriminant function scores by group, for each function shown in Table 9.3.4. With the discriminant functions, the 102 recordings are represented in the three-dimensional space created by the three functions, as shown in Figure 9.3.1 (a), where the first and second functions are displayed. The results show that the first discriminant function mainly discriminates the birdsongs from the other three categories; the second function mainly discriminates the water sounds; the third function mainly discriminates the wind sounds and urban sounds. Birdsongs have high pitch values; water sounds have high slopes (about 0) of spectrum density in the range of [0.005-0.1Hz], i.e. quick variation in the range of [0.005-0.1Hz]; wind sounds have low tonality, while urban sounds have high tonality.



a b

Figure 9.3.1 Plots of the four categories of sound with discriminant functions (a) based on psychoacoustic, music and 1/f noise indices, and (b) based on the indices with fountain sounds labelled as water sounds

The predicted classification results are shown in Table 9.3.5, in terms of the number and percentage of cases correctly and incorrectly classified both by the original functions and in cross validations. For the original functions, it can be seen that all the water and birdsong cases are correctly classified. One of the wind cases is incorrectly classified to the water group. One of the urban cases is incorrectly classified to the water group. The percentage of correctly classified cases of the four categories are 100% for water and bird, 95.7% for wind, and 94.1% for urban. These results are tested through cross validations, in which each case is classified by the functions derived from all cases other than that one. The accuracies in validations are a little lower than those by the original functions for water, wind and bird, and 52.9% for urban. Overall for the four categories, 98.0% of originally grouped cases and 86.3% of cross-validated grouped cases are correctly classified. It suggests the accuracies of the prediction are generally good for the categories of water, wind and bird, but a bit lower for urban.

9.3.2 Discriminant function analysis based on the psychoacoustic, music, and 1/f noise indices with fountain sounds labelled as water sounds

Since the relatively low accuracy of the prediction of urban sounds may result from the various sound types in urban sound category, additional model, or discriminant functions, is developed, in which the fountain sounds in urban category are labelled as in water sound category. Based on the 32 indices as used above, similarly, three canonical linear discriminant functions are developed. Table 9.3.6 shows the three discriminant functions have great correlations with the groupings, with the correlation coefficients larger than

0.8, which indicates that the functions are effective for the discriminating. Also, all the three functions show discriminating ability in the tests of the null hypothesis that a function, and all functions that follow, have no discriminating ability, since the p-values (Sig.) associated with the Chi-square statistic are smaller than the alpha level of 0.05 for all the three tests, which rejects the null hypothesis.

Table 9.3.6 Eigenvalues and Wilks' Lambda of discriminant functions based on psychoacoustic, music, and 1/f noise indices with fountain sounds labelled as water sounds

Eigenvalues					Wilks' Lambda			
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation	Test of Function(s)	Wilks' Lambda	Chi-square	Sig.
1	24.851	75.2	75.2	.980	1 through 3	.002	534.419	.000
2	5.419	16.4	91.6	.919	2 through 3	.041	264.474	.000
3	2.770	8.4	100.0	.857	3	.265	110.155	.000

Structure matrix of the discriminant functions, shown in Table 9.3.7, displays the correlations between the discriminating variables and the dimensions created with the discriminant functions. Somehow similarly to those of the last model, the first function has relatively high correlations (larger than 0.3) with pitch values, including PV AVE and PV1. The second function has relatively high correlations (about 0.3) with 1/f noise and variation of tonality, including S slope [0.005-0.1Hz], Ton slope [0.005-10Hz] and Ton STDEV. The third function also has relatively high correlations (larger than 0.3) with 1/f noise and variation of tonality, including N slope [0.005-0.1Hz], Ton slope [0.005-10Hz], N slope [0.005-10Hz] and Ton STDEV.

The predicted classification results are shown in Table 9.3.5. It can be seen that, for the original functions, same as those of the last model, all the water and birdsong cases are correctly classified. One of the wind cases is incorrectly classified to the water group. One of the urban cases is incorrectly classified to the water group again. In cross validations, percentages of correctly classified cases are lower than those by the original functions, but higher or the same as those in cross validations of the last model. Overall for the four categories, 98.0% of originally grouped cases and 90.2% of cross-validated grouped cases are correctly classified. The relatively low accuracies in cross validations of both the models suggest that to validate the accuracies by the original discriminant functions, more cases may be needed in the models.

Comparing the predicted classification results based on all the psychoacoustic, music, and 1/f noise indices and those based on the psychoacoustic and music indices together, and those based on psychoacoustic indices or on music indices in Chapter 7 Section 7.4.3, it can be seen that the proportions of correctly classified cases based on all the psychoacoustic, music, and 1/f noise indices are higher than the others for all the four

categories, which are above or about 90% for the four sound categories with original functions and for the three natural sound categories in cross validations, and about 50% for the urban sound category in cross validation. These results suggest the prediction accuracy based on all the indices is good, except for urban category that larger sample size may be needed.

Table 9.3.7 Structure matrix of discriminant functions based on psychoacoustic, music, and 1/f noise indices with fountain sounds labelled as water sounds

	Function 1	Function 2	Function 3
PV AVE	.503*	-.210	.011
PV1	.303*	-.107	.019
S STDEV	.290*	-.058	-.082
PA1	.288*	.051	-.220
PA AVE	.252*	.034	-.118
AS AVE	.229*	.044	-.183
PN	-.204*	-.143	.047
ED(E)	.161*	.050	-.055
R AVE	-.106*	.023	-.058
PV slope [0.05-10Hz]	-.101*	-.093	.076
S slope [0.005-0.1Hz]	.037	-.288*	-.284
SF AVE	.021	.223*	-.170
L AVE	-.077	.188*	-.044
S AVE	.119	-.144*	-.105
N STDEV	.017	.144*	-.134
N AVE	-.090	.124*	-.056
N slope [0.005-0.1Hz]	.056	-.238	-.366*
Ton slope [0.005-10Hz]	-.174	-.290	.351*
N slope [0.005-10Hz]	.064	-.232	-.350*
Ton STDEV	.074	.289	-.316*
Ton AVE	.047	.257	-.297*
N slope [1-10Hz]	-.074	-.032	.234*
PA slope [0.05-10Hz]	-.093	-.153	.218*
Fls AVE	.203	.074	-.211*
N slope [0.1-1Hz]	.036	-.006	-.207*
ED(SF)	-.064	.037	.200*
IOI(E) AVE	-.095	.008	.197*
Fls STDEV	.184	.088	-.196*
BS	.134	.117	-.180*
R STDEV	.045	.074	-.151*
L STDEV	.113	.011	-.130*
IOI(SF) AVE	.018	-.024	-.121*

9.4 Automatically Identification of Sound Categories with ANN Based on the Key Indices

In this section, artificial neural network (ANN) models are explored to automatically identify sound categories based on the psychoacoustic and music parameters and 1/f noise behaviour of the parameters, in addition to the discriminant function analyses (DFAs) in Section 9.3. In terms of the key 1/f noise indices as discussed in Section 9.1, since the

1159 30-second segments of the recordings are used as cases for ANNs in this section, the 1/f noise behaviours are calculated over the frequency range of 0.05-10Hz. Thus, the key 1/f noise indices used in this section include N slope [0.05-10Hz], N slope [1-10Hz], N slope [0.1-1Hz], Ton slope [0.05-10Hz], PV slope [0.05-10Hz], and PA slope [0.05-10Hz].

Table 9.4.1 Network design and training information, and statistics results of Networks 1-4

Network				Network 1	Network 2	Network 3	Network 4
Input data				AVE, STD of L, N, S, Ton, R, Fls; PV1, PA1, PN, PV AVE, PA AVE, IOI(E) AVE, ED(E), AS AVE, BS	AVE, STD of L, N, S, Ton, R, Fls; PV1, PA1, PN, PV AVE, PA AVE, IOI(E) AVE, ED(E), AS AVE, BS, N slope [0.05-10Hz], N slope [1-10Hz], N slope [0.1-1Hz], Ton slope [0.05-10Hz], PV slope [0.05-10Hz], PA slope [0.05-10Hz]	AVE, STD of L, N, S, Ton, R, Fls; PV1, PA1, PN, PV AVE, PA AVE, IOI(E) AVE, ED(E), AS AVE, BS	AVE, STD of L, N, S, Ton, R, Fls; PV1, PA1, PN, PV AVE, PA AVE, IOI(E) AVE, ED(E), AS AVE, BS, N slope [0.05-10Hz], N slope [1-10Hz], N slope [0.1-1Hz], Ton slope [0.05-10Hz], PV slope [0.05-10Hz], PA slope [0.05-10Hz]
Network Architecture	Number of Layers			3	3	3	3
	Nodes of Input Layer			21	27	21	27
	Nodes of Hidden Layer			8	8	8	8
	Nodes of Output Layer			4	4	4	4
Training Information	Iterations			10000	10000	10000	10000
	Learn Rate			0.0100	0.0100	0.0100	0.0100
	Momentum Factor			0.8	0.8	0.8	0.8
	Fast-Prop Coef			0.0	0.0	0.0	0.0
	Training Error			0.046	0.052	0.037	0.042
	Test Set Error			0.088	0.117	0.126	0.111
Statistics	Training	Correlation	Node 1	0.985	0.989	0.993	0.993
			Node 2	0.999	0.995	0.995	0.990
			Node 3	1.000	1.000	1.000	1.000
			Node 4	0.991	0.964	0.998	0.998
		Std Dev	Node 1	0.082	0.068	0.056	0.056
			Node 2	0.023	0.042	0.044	0.060
			Node 3	0.005	0.004	0.006	0.009
			Node 4	0.035	0.065	0.015	0.015
	Test	Correlation	Node 1	0.956	0.985	0.932	0.948
			Node 2	0.973	0.958	0.928	0.937
			Node 3	1.000	0.985	0.995	1.000
			Node 4	0.985	0.853	0.963	0.980
		Std Dev	Node 1	0.139	0.081	0.173	0.153
			Node 2	0.099	0.126	0.167	0.156
			Node 3	0.001	0.083	0.045	0.001
			Node 4	0.043	0.159	0.065	0.045

Two networks are developed firstly in this section. Network 1 is based on 21 indices as inputs, i.e., the key psychoacoustic and music indices as discussed in Section 9.1, exclusive of the rhythm indices based on spectral flux method, which do not show any significant difference among the four sound categories. Network 2 is based on 27 indices as inputs, which includes the key psychoacoustic and music indices used in Network 1 and additionally the 1/f noise indices. For each network, the number of input nodes is

equal to the number of input indices used in the model. Four output nodes are designed for each of the networks, as the recordings are to be identified into four categories: water, wind, bird and urban sounds. One hidden layer is chosen for the two networks. The number of nodes in hidden layer is optimized by adjusting the number and examining the nodes' percentage contributions to output signals of that layer, which show the degree of hidden nodes being utilized by the network. The network inputs are normalized automatically between 0 and 1 in order to improve training characteristics; the output, as in a binary form, does not require normalization. The detailed design information of the networks, including input data, network structures, and part of the settings for training parameters, are shown in Table 9.4.1, as well as prediction results.

Apart from the training set, test sets of data are not used to train the network, but monitor the network's responses to cases outside the training set, to reflect network's overtraining status and to check the integrity of model. In this section, the 30-second segments of the recordings are used as cases for training and testing the neural networks. For both Networks 1 and 2, 1159 cases are used, among which 159 cases are selected randomly for testing, with the remaining 1000 cases used for training. Complete histories of root-mean-square (RMS) error between the network's output response and the training targets monitored during training for both training and test set errors of both the networks show that they are generally at an appropriate level of training and are convergent.

From the prediction results shown in Table 9.4.1, it can be seen, for both the networks, that the correlations between network predictions and targets are high, all above 0.95 for the four categories of both training and test sets, except for Output Node 4 (for identifying urban sound category) of Network 2. These results generally suggest the good prediction abilities of the two networks. To examine the details of the prediction errors, the cases with relatively large errors are analysed in Table 9.4.2. The same as those in Chapter 4 Section 4.3, Group 1 includes cases that the difference between network prediction and target is over 0.4 in any of the four output nodes. Among these cases, Group 2 includes cases that the output node with the highest value among the four does not match the target node and thus incorrectly classified. For Network 1, cases in Group 2 show that one wind sound case is classified as water sound; all the five fountain cases and three machine sound cases (sprinkler) in the urban sound category are classified into water sound category; while all the water sounds and birdsongs are identified correctly. For Network 2, one wave sound in the water category is classified as wind sound; one wind sounds are classified as urban sound; in the urban sound category, four fountains are classified into water sound category, two traffic sound is classified as wind sound, and one traffic sound and one voice are classified as birdsongs. Between the two networks, the prediction performances do not differ greatly, based on both the proportions

of accurately predicted cases outside Group 1 (i.e. cases that the differences between network prediction and target are below 0.4) and the proportions of correctly predicted cases outside Group 2. For both networks, the prediction accuracies, according both to Groups 1 and 2, are above about 99% for the three natural sound categories and about 90% for the urban sound category.

Comparing the accuracies of these two networks with those of the former eight networks in Chapter 4 Section 4.3, it can be seen that two networks in this section, which are based on the key psychoacoustic and music indices and on the key psychoacoustic, music, and 1/f noise indices have better performance than the eight networks in Chapter 4, which are based on the psychoacoustic indices only. It suggests that generally the key psychoacoustic and music indices together (or the key psychoacoustic, music, and 1/f noise indices together) have good prediction ability in automatic identification of sound categories with ANN.

Table 9.4.2 Detailed prediction errors of Networks 1-4. For cases in Groups 1 and 2, ratios of the number of cases with high prediction error to the total number of cases in each of the four sound categories are displayed.

	Category	Predicted as	Network 1	Network 2	Network 3	Network 4
Group 1	Water	-	0/367 (100.0%)	2/367 (99.5%)	6/372 (98.4%)	3/372 (99.2%)
	Wind	-	3/283 (98.9%)	1/283 (99.6%)	4/283 (98.6%)	6/283 (97.9%)
	Birdsong	-	0/429 (100.0%)	0/429 (100.0%)	3/429 (99.3%)	0/429 (100.0%)
	Urban	-	9/80 (88.8%)	9/80 (88.8%)	0/75 (100.0%)	0/75 (100.0%)
Group 2	Water	Wind	-	1 wave into cove	2 wave on shingle, 1 wave on sand, 1 wave into cove	2 wave on shingle, 1 wave on sand
		Birdsong	-	-	-	-
		Urban	-	-	-	-
		Total	0/367 (100.0%)	1/367 (99.7%)	4/372 (98.9%)	3/372 (99.2%)
	Wind	Water	1 wind in heath	-	1 wind in deciduous trees, 2 wind in heath	1 wind in deciduous trees, 2 wind in coniferous trees, 2 wind in heath
		Birdsong	-	-	-	-
		Urban	-	1 wind in coniferous trees	-	1 wind in coniferous trees
		Total	1/283 (99.6%)	1/283 (99.6%)	3/283 (98.9%)	6/283 (97.9%)
	Birdsong	Water	-	-	-	-
		Wind	-	-	-	-
		Urban	-	-	-	-
		Total	0/429 (100.0%)	0/429 (100.0%)	0/429 (100.0%)	0/429 (100.0%)
	Urban	Water	5 fountain, 3 machine	4 fountain	-	-
		Wind	-	2 traffic	-	-
		Birdsong	-	1 traffic, 1 voice	-	-
		Total	8/80 (90.0%)	8/80 (90.0%)	0/80 (100.0%)	0/75 (100.0%)

Two additional networks are developed, in which the fountain sounds in urban category are labelled as water sound. Networks 3 and 4 are respectively based on the same indices as in Networks 1 and 2. The detailed design information of the networks, including input data, network structures, part of the settings for training parameters, and prediction results are shown in Table 9.4.1. It can be seen that the correlations between network predictions and targets are high for both the networks, above 0.99 of the four categories for training and above about 0.93 for test. The details of the prediction errors are shown in Table 9.4.2. For both Networks 3 and 4, cases in Group 2 show that a few numbers of water and wind sound cases are classified into the each other's category; and all the birdsongs and urban sounds are identified correctly. Between the two networks, based on both the proportions of accurately predicted cases outside Group 1 and the proportions of correctly predicted cases outside Group 2, the prediction performances are similar. The prediction accuracies, according both to Groups 1 and 2, are above about 98% for all the four sound categories for both networks. Comparing with Networks 1 and 2, it shows that when fountain sounds in urban category are labelled as water sound, the accuracies of the urban category become much higher, although the accuracies of the water and wind categories become a little lower than those when fountain sounds are labelled as urban sound. These results generally suggest the very good prediction abilities of the ANNs, with accuracies over about 98% for the four sound categories.

In the previous research of environmental sound recognition, for identification based on single source recordings, the classification rate reached 70% using the combination of continuous wavelet transform or mel frequency cepstral coefficients (MFCCs) as feature extraction with dynamic time warping as classification (Cowling and Sitte 2003). For real-world environmental sounds, Bunting et al. (2009) used time-domain signal coding (TDSC) combined with learning vector quantization (LVQ) network after a source separation algorithm. However, the accuracy was not high. Aucouturier et al. (2007; Aucouturier and Defreville 2007) used the long-term statistical distribution of frame-based MFCC vectors and Gaussian mixture models (GMMs), and proved precision of 0.9 in the first five nearest neighbours. Different from all these methods, this study uses psychoacoustic/music and 1/f noise features combined with ANNs, and achieves pretty high prediction accuracies. However, the accuracies are not directly comparable across the studies, since the statistic methods of accuracy differ, as well as the sound samples.

9.5 Conclusions

Principal components of all indices, including mean, standard deviation, maximum, minimum, percentiles, and 1/f noise of the psychoacoustic parameters (both psychoacoustic parameters used in Chapter 4 and music features in Chapter 7) are analysed. The main dimensions of the indices are fluctuation and rhythm properties and pitch properties, loudness properties, and 1/f noise properties.

The characteristics of the different categories of sound are shown in terms of a number of key psychoacoustic, music, and 1/f noise indices, by summarising the results in Chapters 4, 7 and 8. Generally, water sounds have a wide range of loudness, low pitch value and strength, low fluctuation strength, low event density (high event interval), low attack slope of event, and do not show periodicity. Wind sounds have a wide range of loudness, low sharpness, low pitch value and strength, low fluctuation strength, low event density (high event interval), low attack slope of event, and do not show periodicity. Birdsongs have low loudness, high sharpness, high pitch value and strength, high fluctuation strength, high event density (low event interval), high attack slope of event, and may show periodicity. Birdsongs exhibit generally 1/f noise behaviour in short and medium time intervals. Urban sounds have high loudness, low pitch values and a relatively wide range of pitch strengths, a relatively wide range of fluctuation strength, event density, and attack slope of event, and may show periodicity. The 1/f noise behaviours of urban sounds also vary in a relatively wide range.

Discriminant functions are developed to automatically identify the sound category of recordings, with all the psychoacoustic, music, and 1/f noise indices. The percentage of correctly identified cases is generally above about 95% for the four categories, but much lower in validation test. The first discriminant function is mainly impacted by pitch value and discriminates the birdsongs from the other three categories; the second function is mainly impacted by slope for 1/f noise behaviour and discriminates the water sounds; and the third functions is mainly impacted by tonality and discriminates between the wind sounds and urban sounds. It suggests that pitch, slope for 1/f noise behaviour and tonality have the greatest ability in discriminating the categories.

With ANNs based on all the psychoacoustic, music, and 1/f noise indices, the prediction accuracies are above about 99% for the three natural sound categories and about 90% for the urban sound category. When fountain sounds in urban category are labelled as water sound, the accuracies are over about 98% for the four sound categories. These results generally suggest the very good prediction abilities of the ANNs.

Chapter 10
Conclusions and future works

Among various sounds in the environment, natural sounds, such as water sounds and birdsongs, have proven to be highly preferred by humans, but the reasons for these preferences have not been thoroughly researched. This study explores differences between various natural and urban environmental sounds from the viewpoint of objective measures in three aspects, which are psychoacoustic parameters that have been recommended in previous soundscape research, additional psychoacoustically related parameters that have mainly been applied in music perception, and 1/f noise that related to dynamic. By analysing recordings of single sound source of categories of water, wind, birdsongs, and urban sounds including street music, mechanical sounds and traffic noise, this study shows a series of differences among different sound types with each aspect of the objective measures, based on a number of statistic methods including one-way AVONA, hierarchical cluster and principal components analyses. The main dimensions of the objective measures are shown. Discriminant functions and artificial neural networks are made to automatically identify the sound categories.

10.1 Contributions of This Study

The results of this study contribute to a number of aspects in soundscape research. First, a number of objective measures that related to subjective hearing sensation and have been used in music are studied and applied to soundscape study. It is expected that these music features would be helpful in the future soundscape research. Second, a series of differences among different sound types are shown with the objective measures. From another point of view, these results also provide reference values in the objective measures with the common types of natural and urban sound in soundscapes. Third, the important factors of the objective measures are shown. Finally, sound categories are automatically identified with models. As the sound type information has important impact on the soundscape evaluation, it is expected the automatically identified sound type information would provide an additional, important dimension of factors for soundscape evaluation simulation or prediction.

This research study has answered the initial questions with that the differences do exist between natural and urban sounds, and among various sound types in terms of the objective measures. It demonstrates the relationship between the objective measures and sound types. As a result, both the objective measures, including acoustic and psychoacoustic properties, and sound types have possibilities to affect physiological or psychological response of human to soundscape, e.g., health and emotion, and to influence the evaluation of soundscape through perception and cognition.

10.1.1 Application of music features in soundscape

Besides the conventional parameters, e.g., A-weighted sound pressure level, more possible parameters are explored for soundscape measurement from music features that have been used for music perception and cognition. From them, pitch and rhythm features are found their appliablity to environmental sounds, both of which are psychoacoustically related. Based on the algorithms of the music features proposed in literature, a number of models are implemented, selected, and simplified/modified for environmental sounds. They are pitch model according to temporal theories, event detection models based on overall envelope and on spectral flux, and periodicity model with the method of beat spectrum. A number of descriptors or parameters are derived from these models, which are pitch value, pitch strength, percentage of audible pitches over time, event interval, event density, attack slope (or spectral flux), and periodicity. A series of statistic indices are thus developed to describe the parameters with time.

The correlations between the pitch and rhythm indices and the loudness and timbre indices indicate that the pitch and rhythm indices developed provide additional variance to the previous psychoacoustic indices that had been analysed with in soundscape studies, since in general the correlations between the two sets of indices are not high, although there are certain correlations between, e.g., pitch and sharpness, pitch strength and tonality, attack slope and fluctuation, and spectral flux and SPL or loudness.

10.1.2 Characteristics of different types of sound

In terms of psychoacoustic parameters, water sounds have low fluctuation strength and a wide range of loudness; wind sounds have low fluctuation strength, a wide range of loudness and low sharpness; birdsongs have high fluctuation strength, high sharpness and low loudness; and urban sounds have high loudness. In terms of the differences between natural and urban sounds, generally speaking, urban sounds have high fluctuation strength

and loudness, while natural sounds have either low fluctuation strength and varied loudness, or high fluctuation strength and low loudness.

For music features, in terms of pitch, both water and wind sounds have relatively low pitch values and low pitch strengths; birdsongs have relatively high pitch values and high pitch strength; and urban sounds generally have low pitch values and a relatively wide range of pitch strengths. The numbers of audible pitches over the duration of water and wind sounds are higher than urban, and than birdsongs. In terms of rhythm, water and wind sounds both have relatively high event interval, low event density, and low attack slope of event based on the envelope method; oppositely, birdsongs have relatively low event interval, high event density, and high attack slope of event; and urban sounds have a relatively wide range of event interval, event density and attack slope. A number of birdsongs and urban sounds show periodicity, while water and wind sounds do not.

The 1/f noise behaviour can be seen as a statastic index in addition to those used for the above analyses in terms of psychoacoustic and music parameters, such as mean and standard deviation. With the 1/f noise behaviour of the results of the psychoacoustic and music parameters, it shows that stream and river sounds in water category exhibit quick variations in loudness and sharpness in the full time interval of 0.1s to 200s. Sea waves sounds in water category exhibit slow variations in loudness and sharpness in short and medium time intervals, i.e., 0.1s to 10s, and exhibit quick variations in long time interval of 10s to 200s. Wind sounds exhibit quick variations in loudness and sharpness in short time interval of 0.1s to 1s, and slow variations in medium and long time intervals. Birdsongs exhibit generally 1/f noise behaviour in short and medium time intervals, and quick variations in loudness and sharpness in long time interval. For pitch, water and wind sounds exhibit slightly quicker variations in pitch than birdsongs, all quicker than 1/f noise. The urban sounds vary in a relatively wide range in terms of 1/f noise, and are generally mixed with many of the sounds in the other three categories.

10.1.3 Main dimensions of the measures

The main dimensions or principal components of the objective measures, including all the psychoacoustic, music, and 1/f noise indices, are examined. For the psychoacoustic indices, which include statistic indices of loudness and timbre, the first component mainly represents loudness proprieties, including average and minimum indices of level, loudness and roughness; the second component mainly represents fluctuation proprieties, including average, standard deviation, maximum, and minimum of fluctuation strength, and maximum of level, loudness, sharpness, and roughness; the third component mainly

represents average indices of sharpness and tonality. From the principal components, three key indices are identified: They are average values of loudness, fluctuation strength, and sharpness. For the music indices, i.e., pitch and rhythm, the principal components analyses suggest that the statistic indices of each parameter, such as average, stand deviation, and percentiles, mostly contribute to one single dimension. The first component mainly represents pitch properties, i.e., pitch value, pitch strength, percentage of audible pitches over time, and rhythm properties of event interval, event density, and attack slope based on envelope method; the second component mainly represents rhythm properties of event interval and event density based on spectral flux method; and the third component mainly represents spectral flux. For all the psychoacoustic, music, and 1/f noise indices, the main dimensions are fluctuation and rhythm properties and pitch properties, loudness properties, and 1/f noise properties, based on the sound recordings used in this study.

10.1.4 Automatic identification of sound types

Discriminant functions and artificial neural networks are developed to automatically identify the sound category of recordings. With all the psychoacoustic, music, and 1/f noise indices, the percentage of correctly identified cases is generally above about 95% for the four categories, i.e., water, wind, bird, and urban, by discriminant functions, but much lower in validation test. Pitch, 1/f noise behaviour, and tonality have the greatest ability in discriminating the categories.

The ANNs have better performance than the discriminant functions for the classification. With ANNs based on all the psychoacoustic, music, and 1/f noise indices, the prediction accuracies are above about 99% for the three natural sound categories and about 90% for the urban sound category. When fountain sounds in urban category are labelled as water sound, the accuracies are over about 98% for the four sound categories. Without the influence of loudness, the accuracies based on only timbre indices are still above 80% for the natural sound categories and 57% for the urban sound category.

10.2 Future Works

The results of this research provide a fundamental for the further soundscape studies. Since both the objective properties and sound type information have the possibility to influence the evaluation of soundscape as indicated by this study, to further study the reason for human's preference to natural sounds, it would be helpful to further link the

objective measures with emotional evaluation of soundscapes. That is, if clear link existed between the two, objective properties of sound would very likely have psychological effect on the soundscape evaluation and preference. The success in the field of psychology of music in terms of the links between music features and music-raised emotions provides a cue for the study of link between the psychoacoustic parameters and emotional responses in soundscapes. Much knowledge and methodology could be borrowed from music to the research of subjective emotional evaluation in soundscapes.

Furthermore, the influence of objective properties of environmental sound to people's evaluation of emotion and preference, if existed, could be verified and studied in subjective evaluation test in laboratory condition by controlling the values of objective measures with modifications of the sound samples through computer processing, e.g., by sound filtering with attenuation or intensification of audio power in certain frequency bands. Consequently, according to the results, soundscape design could be researched in terms of the possibility of transferring the mode of modifications into practice in real environment. For instance in the way that, to obtain the similar effect of those in the computer processing, introducing landscape elements created by vegetation or structures to environment, which were designed to modify the sound spectrum distribution through propagation.

In this study, the pitch and rhythm algorithms for environment sound have been obtained based on the systematic review of music features proposed in literature for perception of the auditory system, and on the applicability and feasibility through examination of their performances with common types of sound in soundscapes. Also, the psychoacoustic parameters, including loudness and timbre, were calculated with a number of existing psychoacoustic algorithmic models that have been well developed based on physical stimuli. For meaningful environmental sounds, the correlation of the parameters (including all loudness, pitch, timbre, and rhythm) calculated by these methods with the subjective sensations may be verified in future works through subjective listening tests.

While this study focuses on the characteristics and identifications of single source sounds in soundscapes, their applications could be further put in the context of soundscapes with mixed sound sources. As most soundscape environments contain various sound source components – although simplifying assumption could be made that at any short time interval soundscape is dominated by a particular sound source, it is expected that the evaluation of soundscape would be more complex. However, the results of this study would benefit further studies of soundscapes with multiple sound sources as a basis in their characteristics and identifications. Also, the influence of environments on

soundscape characteristics, such as reverberation in urban spaces, could be taken into consideration.

References

Alonso, M., David, B. and Richard, G. (**2004**). "Tempo and beat estimation of musical signals." in Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004), Barcelona, Spain, 158-163.

Alvarsson, J. J., Wiens, S. and Nilsson, M. E. (**2010**). "Stress recovery during exposure to nature sound and environmental noise." International Journal of Environmental Research and Public Health **7**(3), 1036-1046.

ASA (**1960**). *Acoustical terminology SI, 1-1960*. American Standards Association, New York, US.

Ashmore, J. F. (**1987**). "A fast motile response in guinea-pig outer hair cells: The cellular basis of the cochlear amplifier." The Journal of Physiology **388**, 323-347.

Assmann, P. F. and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies." The Journal of the Acoustical Society of America **88**(2), 680-697.

Aucouturier, J.-J. and Defreville, B. (**2007**). "Sounds like a park: A computational technique to recognize soundscapes holistically, without source identification." in Proceedings of 19th International Congress on Acoustics (ICA 2007), Madrid, Spain.

Aucouturier, J.-J., Defreville, B. and Pachet, F. (**2007**). "The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music." The Journal of the Acoustical Society of America **122**(2), 881-891.

Aures, W. (**1985**). "Berechnungsverfahren fur den sensorischen wohlklang beliebiger schallsignale." ("A model for calculating the sensory euphony of various sounds") Acustica **59**, 130-141.

Axelsson, O., Nilsson, M. E. and Berglund, B. (**2010**). "A principal components model of soundscape perception." The Journal of the Acoustical Society of America **128**(5), 2836-2846.

Bak, P., Tang, C. and Wiesenfeld, K. (**1987**). "Self-organized criticality: An explanation of the 1/f noise." Physical Review Letters **59**(4), 381-384.

Balkwill, L. L. and Thompson, W. F. (**1999**). "A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues." Music Perception **17**(1), 43-64.

Behringer (**2013**). http://www.behringer.com/EN/Products/B-2-PRO.aspx/ Last accessed on 17 Jul 2013.

Bello, J. P., Duxbury, C., Davies, M. and Sandler, M. (**2004**). "On the use of phase and energy for musical onset detection in complex domain." IEEE Signal Processing Letters **11**, 553–556.

Benfield, J. A., Bell, P. A., Troup, L. J. and Soderstrom, N. (**2010**). "Does anthropogenic noise in national parks impair memory?" Environment and Behavior **42**(5), 693-706.

Berglund, B., Harder, K. and Preis, A. (**1994**). "Annoyance perception of sound and information extraction." The Journal of the Acoustical Society of America **95**(3), 1501-1509.

Berglund, B. and Nilsson, M. E. (**2006**). "On a tool for measuring soundscape quality in urban residential areas." Acta Acustica united with Acustica **92**(6), 938-944.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J. and Dacquet, A. (**2005**). "Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts." Cognition & Emotion **19**(8), 1113-1139.

Blenheim Palace (**2013**). http://www.blenheimpalace.com/ Last accessed on 17 Jul 2013.

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

Borgatta, E. F. (**1961**). "Mood, personality, and interaction." Journal of General Psychology **64**(1), 105-137.

Botteldooren, D., De Coensel, B. and De Muer, T. (**2006**). "The temporal structure of urban soundscapes." Journal of Sound and Vibration **292**(1-2), 105-123.

Botteldooren, D. and Verkeyn, A. (**2002**). "Fuzzy models for accumulation of reported community noise annoyance from combined sources." The Journal of the Acoustical Society of America **112**(4), 1496-1508.

Brandenburg, K. (**1999**). "MP3 and AAC explained." in Proceedings of Audio Engineering Society (AES) 17th International Conference on High Quality Audio Coding, Florence, Italy, 139-146.

British Library Sound Archive (**2013a**). http://www.bl.uk/reshelp/bldept/soundarch/index.html/ Last accessed on 17 Jul 2013.

British Library Sound Archive (**2013b**). http://sounds.bl.uk/ Last accessed on 17 Jul 2013.

Brown, A. L., Kang, J. and Gjestland, T. (**2011**). "Towards standardization in soundscape preference assessment." Applied Acoustics **72**(6), 387-392.

Brown, J. C. (**1993**). "Determination of the meter of musical scores by autocorrelation." The Journal of the Acoustical Society of America **94**(4), 1953-1957.

Bunting, O., Stammers, J., Chesmore, D., Bouzid, O., Tian, G. Y., Karatsovis, C. and Dyne, S. (**2009**). "Instrument for soundscape recognition, identification and evaluation (ISRIE): Technology and practical uses." in Proceedings of Euronoise 2009, Edinburgh, Scotland.

Burns, E. M. and Ward, W. D. (**1982**). "Intervals, scales, and tuning." *The Psychology of Music*. D. Deutsch (ed). Academic Press, New York, 241-269.

Cain, R., Jennings, P. and Poxon, J. (**2013**). "The development and application of the emotional dimensions of a soundscape." Applied Acoustics **74**(2), 232-239.

Cardinal, R. (**2013**). "Pairwise comparisons in sas and spss." available at http://egret.psychol.cam.ac.uk/statistics/local_copies_of_sources_Cardinal_and_Aitken_ANOVA/MultipleComparisons_3.htm - b20/ Last accessed on 20 Jul 2013.

Carles, J. L., Barrio, I. L. and de Lucio, J. V. (**1999**). "Sound influence on landscape values." Landscape and Urban Planning **43**(4), 191-200.

Cowling, M. and Sitte, R. (**2003**). "Comparison of techniques for environmental sound recognition." Pattern Recognition Letters **24**(15), 2895-2907.

Crowder, R. G. (**1985**). "Perception of the major/minor distinction: III. Hedonic, musical, and affective discriminations." Bulletin of the Psychonomic Society **23**(4), 314-316.

Cusack, P. (**2001**). *Your Favourite London Sounds.* CD-ROM, London Musicians' Collective, RESFLS1CD.

Davies, W. J., Adams, M. D., Bruce, N. S., Cain, R., Carlyle, A., Cusack, P., Hume, K. I., Jennings, P. and Plack, C. J. (**2007**). "The positive soundscape project." in Proceedings of the 19th International Congress on Acoustics (ICA 2007), Madrid, Spain, paper ENV-10-004.

de Boer, E. (**1977**). "Pitch theories unified." *Psychophysics and physiology of hearing*. E. F. Evans and J. P. Wilson (eds). Academic, London, 323-334.

De Coensel, B. and Botteldooren, D. (**2006**). "The quiet rural soundscape and how to characterize it." Acta Acustica united with Acustica **92**(6), 887-897.

De Coensel, B., Botteldooren, D. and De Muer, T. (**2003**). "1/f noise in rural and urban soundscapes." Acta Acustica united with Acustica **89**, 287-295.

DIN 45631 (**1991**). *Procedure for calculating loudness level and loudness*. Deutsches Institut für Normung e.V., Berlin, Germany.

DIN 45631/A1 (**2010**). *Calculation of loudness level and loudness from the sound spectrum - Zwicker method - Amendment 1: Calculation of the loudness of time-variant sound; with CD-ROM*. Deutsches Institut für Normung e.V., Berlin, Germany.

DIN 45692 (**2009**). *Measurement technique for the simulation of the auditory sensation of sharpness*. Deutsches Institut für Normung e.V., Berlin, Germany.

Dowling, W. J. and Harwood, D. L. (**1986**). *Music Cognition*. Academic Press, San Diego.

Dubois, D., Guastavino, C. and Raimbault, M. (**2006**). "A cognitive approach to urban soundscapes: Using verbal data to access everyday life auditory categories." Acta Acustica united with Acustica **92**(6), 865-874.

Eerola, T., Lartillot, O. and Toiviainen, P. (**2009**). "Prediction of multidimensional emotional ratings in music from audio using multivariate regression models." in Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009), Kobe, Japan, 621-626.

Elvers, G. C. (**2013**). "Using spss for one way analysis of variance." available at http://academic.udayton.edu/gregelvers/psy216/spss/1wayanova.htm/ Last accessed on 20 Jul 2013.

Fastl, H. (**1977**). "Roughness and temporal masking patterns of sinusoidally amplitude modulated broadband noise." *Psychophysics and Physiology of Hearing*. E. F. Evans and J. P. Wilson (eds). Academic, London, 403-415.

Fastl, H. (**1990**). "The hearing sensation roughness and neuronal responses to am-tones." Hearing Research **46**(3), 293-296.

Fastl, H. (**2005**). "Psychoacoustics and sound quality." *Communication Acoustics*. J. Blauert (ed). Springer, Berlin, 139-162.

Fastl, H. and Stoll, G. (**1979**). "Scaling of pitch strength." Hearing Research **1**(4), 293-301.

Fishman, Y. I., Reser, D. H., Arezzo, J. C. and Steinschneider, M. (**2000**). "Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness." The Journal of the Acoustical Society of America **108**(1), 235-246.

Fletcher, H. (**1934**). "Loudness, pitch and the timbre of musical tones and their relation to the intensity, the frequency and the overtone structure." The Journal of the Acoustical Society of America **6**(2), 59-69.

Fletcher, H. (**1940**). "Auditory patterns." Reviews of Modern Physics **12**(1), 47-65.

Fletcher, H. and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation." The Journal of the Acoustical Society of America **5**(2), 82-108.

Fletcher, H. and Munson, W. A. (**1937**). "Relation between loudness and masking." The Journal of the Acoustical Society of America **9**(1), 1-10.

Foote, J., Cooper, M. and Nam, U. (**2002**). "Audio retrieval by rhythmic similarity." in Proceedings of the 3rd International Society for Music Information Retrieval Conference (ISMIR 2002), Paris, France, 265-266.

Foote, J. T. and Cooper, M. L. (**2003**). "Media segmentation using self-similarity decomposition." in Proceedings of SPIE Storage and Retrieval for Media Databases, San Jose, CA, US, 167-175.

Fostex Company (**2013**). http://www.fostexinternational.com/docs/archive_products/FR-2.shtml/ Last accessed on 17 Jul 2013.

Gabrielsson, A. (**1973**). "Adjective ratings and dimension analyses of auditory rhythm patterns." Scandinavian Journal of Psychology **14**(4), 244-260.

Gabrielsson, A. (**2001**). "Emotions in strong experiences with music." *Music and Emotion Theory and Research*. P. N. Juslin and J. A. Sloboda (eds). Oxford University Press, New York, 431-449.

Gabrielsson, A. and Lindström, E. (**2001**). "The influence of musical structure on emotional expression." *Music and Emotion Theory and Research*. P. N. Juslin and J. A. Sloboda (eds). Oxford University Press, New York, 223-248.

Gagnon, L. and Peretz, I. (**2003**). "Mode and tempo relative contributions to "happy-sad" judgements in equitone melodies." Cognition & Emotion **17**(1), 25-40.

Genuit, K. and Fiebig, A. (**2006**). "Psychoacoustics and its benefit for the soundscape approach." Acta Acustica united with Acustica **92**(6), 952-958.

Glasberg, B. R. and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data." Hearing Research **47**(1-2), 103-138.

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

Gomez, P. and Danuser, B. (**2007**). "Relationships between musical structure and psychophysiological measures of emotion." Emotion **7**(2), 377-387.

Goto, M. and Muraoka, Y. (**1995**). "Music understanding at the beat level: Real-time beat tracking for audio signals." in Proceedings of the International Joint Conferences on Artificial Intelligence (IJCAI-95) Workshop on Computational Auditory Scene Analysis, Montreal, Canada, 68-75.

Guastavino, C. and Katz, B. F. G. (**2004**). "Perceptual evaluation of multi-dimensional spatial audio reproduction." The Journal of the Acoustical Society of America **116**(2), 1105-1115.

Guastavino, C., Katz, B. F. G., Polack, J. D., Levitin, D. J. and Dubois, D. (**2005**). "Ecological validity of soundscape reproduction." Acta Acustica united with Acustica **91**(2), 333-341.

Gundlach, R. H. (**1935**). "Factors determining the characterization of musical phrases." American Journal of Psychology **47**, 624-643.

Hall, D. A., Irwin, A., Edmondson-Jones, M., Phillips, S. and Poxon, J. E. W. (**2013**). "An exploratory evaluation of perceptual, psychoacoustic and acoustical properties of urban soundscapes." Applied Acoustics **74**(2), 248-254.

Han, B.-j., Ho, S., Dannenberg, R. B. and Hwang, E. (**2009**). "SMERS: Music emotion recognition using support vector regression." in Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR 2009), Kobe, Japan, 651-656.

Handel, S. (**1989**). *Listening: An Introduction to the Perception of Auditory Events*. Mit Press, Cambridge, MA.

HEAD acoustics GmbH (**2011a**). http://www.head-acoustics.de/eng/nvh_artemis.htm/ Last accessed on 20 Jul 2013.

HEAD acoustics GmbH (**2011b**). "Head application note - Psychoacoustic analyses I." available at http://www.head-acoustics.de/downloads/eng/application_notes/PsychoacousticAnalysesI_06_11e.pdf/ Last accessed on 20 Jul 2013.

HEAD acoustics GmbH (**2011c**). "Head application note - Psychoacoustic analyses II." available at http://www.head-acoustics.de/downloads/eng/application_notes/PsychoacousticAnalysesII_06_11e.pdf / Last accessed on 20 Jul 2013.

Hevner, K. (**1936**). "Experimental studies of the elements of expression in music." American Journal of Psychology **48**, 246-268.

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

Hilton, A. and Armstrong, R. (**2006**). "Post hoc anova tests." available at http://eprints.aston.ac.uk/9317/1/Statnote_6.pdf/ Last accessed on 20 Jul 2013.

IEC 61672 (**2003**). *Electroacoustics - Sound level meters*. International Electrotechnical Commission, Geneva, Switzerland.

ISO 226 (**2003**). *Acoustics - Normal equal-loudness-level contours*. International organization for standardization, Geneva, Switzerland.

ISO 532 (**1975**). *Acoustics - Method for calculating loudness level.* International organization for standardization, Geneva, Switzerland.

ISO/IEC 11172-3 (**1993**). *Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s - Part 3: Audio*. International Organization for Standardization / International Electrotechnical Commission, Geneva, Switzerland.

ISO/IEC 13818-3 (**1998**). *Information technology - Generic coding of moving pictures and associated audio information - Part 3: Audio*. International Organization for Standardization / International Electrotechnical Commission, Geneva, Switzerland.

Jeon, J. Y., Lee, P. J. and You, J. (**2010a**). "Urban space design based on the perceptual assessment of soundscape." The Journal of the Acoustical Society of America **128**(4), 2369.

Jeon, J. Y., Lee, P. J., You, J. and Kang, J. (**2010b**). "Perceptual assessment of quality of urban soundscapes with combined noise sources and water sounds." The Journal of the Acoustical Society of America **127**(3), 1357-1366.

Jolliffe, I. T. (**2002**). *Principal Component Analysis*. Springer, New York.

Kang, J. (**2001**). "Sound propagation in interconnected urban streets: A parametric study." Environment and Planning B: Planning and Design **28**(2), 281-294.

Kang, J. (**2002**). "Numerical modelling of the sound fields in urban streets with diffusely reflecting boundaries." Journal of Sound and Vibration **258**(5), 793-813.

Kang, J. (**2005**). "Numerical modeling of the sound fields in urban squares." The Journal of the Acoustical Society of America **117**(6), 3695-3706.

Kang, J. (**2006**). *Urban Sound Environment*. Taylor & Francis incorporating Spon, London.

Kang, J. and Zhang, M. (**2010**). "Semantic differential analysis of the soundscape in urban open public spaces." Building and Environment **45**(1), 150-157.

Karlsson, H. (**2000**). "The acoustic environment as a public domain." Soundscape: The Journal of Acoustic Ecology **1**,10-13.

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., Speck, J. A. and Turnbull, D. (**2010**). "Music emotion recognition: A state of the art review." in Proceedings of 11th International Society for Music Information Retrieval Conference (ISMIR 2010), Utrecht, Netherlands, 255-266.

Klapuri, A. (**1999**). "Sound onset detection by applying psychoacoustic knowledge." in Proceedings of IEEE International Conference on the Acoustics, Speech, and Signal Processing (ICASSP 1999) **6**, Phoenix, AZ, US, 3089-3092.

Klapuri, A. P., Eronen, A. J. and Astola, J. T. (**2006**). "Analysis of the meter of acoustic musical signals." IEEE Transactions on Audio Speech and Language Processing **14**(1), 342-355.

Krijnders, J. D., Niessen, M. E. and Andringa, T. C. (**2010**). "Sound event recognition through expectancy-based evaluation of signal-driven hypotheses." Pattern Recognition Letters **31**(12), 1552-1559.

Krumhansl, C. L. (**1996**). "A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas." Music Perception **13**(3), 401-432.

Krumhansl, C. L. (**1997**). "An exploratory study of musical emotions and psychophysiology." Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Experimentale **51**(4), 336-353.

Kulkarni, A. D. (**1994**). *Artificial Neural Network for Image Understanding*. Van Nostrand Reinhold, New York.

Lam, K. C., Brown, A. L., Marafa, L. and Chau, K. C. (**2010**). "Human preference for countryside soundscapes." Acta Acustica united with Acustica **96**(3), 463-471.

Laroche, J. (**2004**). "Efficient tempo and beat tracking in audio recordings." Journal of The Audio Engineering Society **51**(4), 226-233.

Lartillot, O. (**2011**). *MIRtoolbox 1.3.4 User's Manual*. available at https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox/MIRtoolbox%20Users%20Guide%201.3.4/ Last accessed on 20 Jul 2013.

Lartillot, O., Eerola, T., Toiviainen, P. and Fornari, J. (**2008**). "Multi-feature modeling of pulse clarity: Design, validation, and optimization." In Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008), Philadelphia, PA, US, 521-526.

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

Leman, M., Vermeulen, V., De Voogdt, L., Moelants, D. and Lesaffre, M. (**2005**). "Prediction of musical affect using a combination of acoustic structural cues." Journal of New Music Research **34**(1), 39-67.

Liberman, M. C. (**1978**). "Auditory-nerve response from cats raised in a low-noise chamber." The Journal of the Acoustical Society of America **63**(2), 442-455.

Licitra, G., Memoli, G., Botteldooren, D. and De Coensel, B. (**2005**). "Traffic noise and perceived soundscapes: A case study." in Proceedings of Forum Acusticum, Budapest, Hungary, 1875-1880.

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

Lorr, M., Daston, P. and Smith, I. R. (**1967**). "An analysis of mood states." Educational and Psychological Measurement **27**(1), 89-96.

Martin, J. G. (**1972**). "Rhythmic (hierarchical) versus serial structure in speech and other behavior." Psychological Review **79**(6), 487-509.

Meddis, R. and Hewitt, M. J. (**1991a**). "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**(6), 2866-2882.

Meddis, R. and Hewitt, M. J. (**1991b**). "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**(6), 2883-2894.

Meddis, R. and O'Mard, L. (**1997**). "A unitary model of pitch perception." The Journal of the Acoustical Society of America **102**(3). 1811-1820.

Moore, B. C. J. (**1977**). "Effects of relative phase of the components on the pitch of three-component complex tones." *Psychophysics and Physiology of Hearing*. E. F. Evans and J. P. Wilson (eds). Academic Press, London, 349-358.

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*. Academic press, london.

Moore, B. C. J. and Glasberg, B. R. (**1996**). "A revision of Zwicker's loudness model." Acustica United with Acta Acustica **82**, 335-345.

Moore, B. C. J., Glasberg, B. R., Plack, C. J. and Biswas, A. K. (**1988**). "The shape of the ear's temporal window." The Journal of the Acoustical Society of America **83**(3), 1102-1116.

Nilsson, M. E. (**2007**). "A-weighted sound pressure level as an indicator of short-term loudness or annoyance of road-traffic sound." Journal of Sound and Vibration **302**(1-2), 197-207.

Nilsson, M. E. and Berglund, B. (**2006**). "Soundscape quality in suburban green areas and city parks." Acta Acustica united with Acustica **92**(6), 903-911.

Noll, A. M. (**1964**). "Short-time spectrum and "cepstrum" techniques for vocal-pitch detection." The Journal of the Acoustical Society of America **36**(2), 296-302.

Noll, A. M. (**1967**). "Cepstrum pitch determination." The Journal of the Acoustical Society of America **41**(2), 293-309.

Palmer, A. R. and Russell, I. J. (**1986**). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells." Hearing Research **24**(1), 1-15.

Parncutt, R. S., H. (**1994**). "Applying psychoacoustics in composition: "Harmonic" progressions of "nonharmonic" sonorities." Perspectives of New Music **32**, 88-129.

Patterson, D. W. (**1996**). *Artificial Neural Networks Theory and Applications*. Prentice Hall, Singapore.

Patterson, R. D. and Moore, B. C. J. (**1986**). "Auditory filters and excitation patterns as representations of frequency resolution." *Frequency Selectivity in Hearing*. B. C. J. Moore (ed). Academic, London, 123-177.

Patterson, R. D., Nimmo-Smith, I., Holdsworth, J. and Rice, P. (**1988**). "Spiral VOS final report, part A: The auditory filterbank." Contract Report (APU 2341), Cambridge Electronic Design.

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

Payne, S. R., Davies, W. J. and Adams, M. D. (**2009**). *Research into the Practical and Policy Applications of Soundscape Concepts and Techniques in Urban Areas*. Technical Report. Department of Environment, Food and Rural Affairs (DEFRA), London.

Payne, S. R., Patrick, D.-W. and Irvine, K. N. (**2007**). "People's perceptions and classifications of sounds heard in urban parks: Semantics, affect and restoration." in Proceedings of 36th International Congress and Exhibition on Noise Control Engineering (Inter-Noise 2007), Istanbul, Turkey, paper in07_233.

Peltonen, V. T. K., Eronen, A. J., Parviainen, M. P. and Klapuri, A. P. (**2001**). "Recognition of everyday auditory scenes: Potentials, latencies and cues." in Proceedings of the 110th Convention of the Audio Engineering Society, Amsterdam, Netherlands.

Percival, G. and Tzanetakis, G. (**2009**) *Marsyas User Manual*. available at http://marsyas.info/assets/docs/manual/marsyas-user/index.html/ Last accessed on 20 Jul 2013.

Peretz, I., Gagnon, L. and Bouchard, B. (**1998**). "Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage." Cognition **68**(2), 111-141.

Pheasant, R., Horoshenkov, K., Watts, G. and Barrett, B. (**2008**). "The acoustic and visual factors influencing the construction of tranquil space in urban and rural environments tranquil spaces-quiet places?" The Journal of the Acoustical Society of America **123**(3), 1446-1457.

Plomp, R. and Levelt, W. J. M. (**1965**). "Tonal consonance and critical bandwidth." The Journal of the Acoustical Society of America **38**(4), 548-560.

Plomp, R., Pols, L. C. W. and van de Geer, J. P. (**1967**). "Dimensional analysis of vowel spectra." The Journal of the Acoustical Society of America **41**(3), 707-712.

Plomp, R. and Steeneken, H. J. M. (**1968**). "Interference between two simple tones." The Journal of the Acoustical Society of America **43**(4), 883-884.

Pols, L. C. W., van der Kamp, L. J. T. and Plomp, R. (**1969**). "Perceptual and physical space of vowel sounds." The Journal of the Acoustical Society of America **46**(2B), 458-467.

Preis, A. and Golebiewski, R. (**2004**). "Noise annoyance perception as a function of distance from a moving source." Noise Control Engineering Journal **52**(1), 20-25.

Pressnitzer, D. and McAdams, S. (**1999**). "Two phase effects in roughness perception." The Journal of the Acoustical Society of America **105**(5), 2773-2782.

Raimbault, M. and Dubois, D. (**2005**). "Urban soundscapes: Experiences and knowledge." Cities **22**(5), 339-350.

Raimbault, M., Lavandier, C. and Bérengier, M. (**2003**). "Ambient sound assessment of urban environments: Field studies in two french cities." Applied Acoustics **64**(12), 1241-1256.

Rasch, R. A. and Plomp, R. (**1982**). "The perception of musical tones." *The Psychology of Music*. D. Deutsch (ed). Academic Press, New York, 1-24.

Reif, F. (**1965**). *Fundamentals of Statistical snd Thermal Physics*. McGraw-Hill, New York.

Rigg, M. G. (**1940**). "Speed as a determiner of musical mood." Journal of Experimental Psychology **27**(5), 566-571.

Rigg, M. G. (**1964**). "The mood effects of music: A comparison of data from four investigators." Journal of Psychology **58**(2), 427-438.

Ritsma, R. J. (**1962**). "Existence region of the tonal residue. I." The Journal of the Acoustical Society of America **34**(9A), 1224-1229.

Ritsma, R. J. (**1967**). "Frequencies dominant in the perception of the pitch of complex sounds." The Journal of the Acoustical Society of America **42**(1), 191-198.

Roederer, J. G. (**1995**). *The Physics and Psychophysics of Music: An Introduction*. Springer-Verlag, New York.

Russell, J. A. (**1980**). "A circumplex model of affect." Journal of Personality and Social Psychology **39**(6), 1161-1178.

Rychtáriková, M. and Vermeir, G. (**2013**). "Soundscape categorization on the basis of objective acoustical parameters." Applied Acoustics **74**(2), 240-247.

Schafer, R. M. (**1977**). *The Tuning of the World*. Knopf, New York.

Scheirer, E. D. (**1998**). "Tempo and beat analysis of acoustic musical signals." The Journal of the Acoustical Society of America **103**(1), 588-601.

Schouten, J. F. (**1968**). "The perception of timbre." in Proceedings of the 6th International Congress on Acoustics, Tokyo, Japan, 35-44.

Schouten, J. F. (**1970**). "The residue revisited." *Frequency Analysis and Periodicity Detection in Hearing*. R. Plomp and G. F. Smoorenburg (eds). Sijthoff, London, 41-58.

Schouten, J. F., Ritsma, R. J. and Cardozo, B. L. (**1962**). "Pitch of the residue." The Journal of the Acoustical Society of America **34**(9B), 1418-1424.

Schubert, E. (**2004**). "Modeling perceived emotion with continuous musical features." Music Perception **21**(4), 561-585.

Schulte-Fortkamp, B., Brooks, B. M. and Bray, W. R. (**2007**). "Soundscape: An approach to rely on human perception and expertise in the post-modern community noise era." Acoustics Today **3**(1), 7-15.

Schulte-Fortkamp, B. and Fiebig, A. (**2006**). "Soundscape analysis in a residential area: An evaluation of noise and people's mind." Acta Acustica united with Acustica **92**(6), 875-880.

Slaney, M. (**1993**). *An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank.* Apple Computer Technical Report #35. Apple Computer Inc., Cupertino, CA.

Sony Creative Software, Inc. (**2013**). http://www.sonycreativesoftware.com/soundforgesoftware/ Last accessed on 20 Jul 2013.

Sottek, R. (**1994**). "Gehörgerechte rauhigkeitsberechnung." ("Hearing model of roughness calculation") in Proceedings of German Annual Conference on Acoustics (DAGA), Dresden, Germany, 1197-1200.

Sottek, R., Vranken, P. and Kaiser, H.-J. (**1994**). "Anwendung der gehörgerechten rauhigkeitsberechnung." ("Application of hearing model of roughness calculation") in Proceedings of German Annual Conference on Acoustics (DAGA), Dresden, Germany, 1201-1204.

Southworth, M. (**1969**). "The sonic environment of cities." Environment and Behavior **1**, 49-70.

Spss Inc. (**2009**). *PASW Statistics Base 18*. available at http://support.spss.com/ProductsExt/Statistics/Documentation/18/clientindex.html/ Last accessed on 19 Nov 2010.

StatSoft Inc. (**2013**). "Discover which variables discriminate between groups, discriminant function analysis." available at http://www.statsoft.com/textbook/discriminant-function-analysis/ Last accessed on 20 Jul 2013.

Stevens, J. (**1999**). "Post hoc tests in anova." available at http://pages.uoregon.edu/stevensj/posthoc.pdf/ Last accessed on 20 Jul 2013.

Stevens, S. S. (**1957**). "On the psychophysical law." Psychological Review **64**(3), 153-181.

Stevens, S. S. (**1972**). "Perceived level of noise by Mark VII and decibels (E)." The Journal of the Acoustical Society of America **51**(2B), 575-601.

Stockburger, D. W. (**2013**). "Discriminant function analysis." available at http://www.psychstat.missouristate.edu/multibook/mlt03m.html/ Last accessed on 20 Jul 2013.

Terhardt, E. (**1972**). "Zur tonhohenwahrnehmung von klangen." ("Perception of the pitch of complex tones") Acustica **26**, 173-199.

Terhardt, E. (**1974a**). "On the perception of periodic sound fluctuations (roughness)." Acustica **30**(3), 201-213.

Terhardt, E. (**1974b**). "Pitch, consonance, and harmony." The Journal of the Acoustical Society of America **55**(5), 1061-1069.

Terhardt, E. (**1979**). "Calculating virtual pitch." Hearing Research **1**(2), 155-182.

Terhardt, E. and Stoll, G. (**1981**). "Skalierung des wohlklangs (der sensorischen konsonanz) von 17 umweltschallen und untersuchung der beteiligten hörparameter." ("Scaling the sensory pleasantness of 17 environmental sounds and investigation of correlated hearing sensations") Acustica **48**, 247-253.

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

Thayer, R. E. (**1989**). *The Biopsychology of Mood and Arousal*. Oxford Univ. Press, Oxford.

Thompson, E. (**2002**). *The Soundscape of Modernity: Architectural Acoustics and the Culture of Listening in America, 1900-1933*. The MIT Press, Cambridge.

Todd, M. P. M. (**1994**). "The auditory "Primal Sketch": A multiscale model of rhythmic grouping." Journal of New Music Research **23**(1), 25-70.

Tolonen, T. and Karjalainen, M. (**2000**). "A computationally efficient multipitch analysis model." IEEE Transactions on Speech and Audio Processing **8**(6), 708-716.

Truax, B. (**1999**). *Handbook for Acoustic Ecology*. CD-ROM version, Cambridge Street Publishing, CSR-CDR 9901.

Tzanetakis, G. and Cook, P. (**2000**). "Marsyas: A framework for audio analysis." Organised Sound **4**, 169-175.

UCLA Statistical Consulting Group (**2013a**). "Annotated spss output discriminant analysis." available at http://www.ats.ucla.edu/stat/spss/output/SPSS_discrim.htm/ Last accessed on 20 Jul 2013.

UCLA Statistical Consulting Group (**2013b**). "Annotated spss output principal components analysis." available at http://statistics.ats.ucla.edu/stat/spss/output/principal_components.htm/ Last accessed on 20 Jul 2013.

Vesta Services Inc. (**2000**). *Qnet 2000 Manual*. available at http://www.qnetv2k.com/ Last accessed on 5 Feb 2013.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds." The Journal of the Acoustical Society of America **66**(5), 1364-1380.

Viemeister, N. F. and Plack, C. J. (**1993**). "Time analysis." *Human Psychophysics*. W. Yost, A. Popper and R. Fay (eds). Springer-Verlag, New York, 116-154.

Viollin, S. (**2003**). "Two examples of audio-visual interactions in an urban context." in Proceedings of the 5th European Conference on Noise Control (Euro-Noise 2003), Naples, Italy, paper 073.

von Bekesy, G. (**1947**). "The variation of phase along the basilar membrane with sinusoidal vibrations." The Journal of the Acoustical Society of America **19**(3), 452-460.

von Bekesy, G. (**1960**). *Experiments in Hearing*. McGraw-Hill, New York.

von Bismarck, G. (**1974a**). "Sharpness as an attribute of the timbre of steady sounds." Acustica **30**, 159-172.

von Bismarck, G. (**1974b**). "Timbre of steady sounds: A factorial investigation of its verbal attributes." Acustica **30**, 146-159.

von Helmholtz, H. L. F. (**1863**). *Die Lehre von den Tonempfindungen als Physiologische Grundlage für die Theorie der Musik*. Vieweg, Braunscheig. trans. A. J. Ellis, *On the*

Sensations of Tone as A Physiological Basis for the Theory of Music. Longmans & Co., London, 1885. Reprint, Dover Pubs, New York, 1954.

Voss, R. F. (**1979**). "1/f (flicker) noise: A brief review." in Proceedings of 33rd Annual Symposium on Frequency Control, Atlantic City, NJ, US, 40-46.

Voss, R. F. and Clarke, J. (**1978**). "1/f noise in music: Music from 1/f noise." Journal of the Acoustical Society of America **63**, 258-263.

Wakefield Council (**2013**). http://www.wakefield.gov.uk/CultureAndLeisure/ParksAndOpenSpaces/HawPark/default.htm/ Last accessed on 17 Jul 2013.

Watson, K. B. (**1942**). "The nature and measurement of musical meanings." Psychological Monographs **54**(2), 1-43.

Wedin, L. (**1972**). "Multidimensional study of perceptual-emotional qualities in music." Scandinavian Journal of Psychology **13**(4), 241-257.

Wessel, D. (**1979**). "Timbre space as a musical control structure." Computer Music Journal **3**(2), 45-52.

Weston Manor (**2013**). http://themanorweston.com/ Last accessed on 17 Jul 2013.

Wightman, F. L. (**1973**). "The pattern-transformation model of pitch." The Journal of the Acoustical Society of America **54**(2), 407-416.

Wikipedia (**2013**). http://en.wikipedia.org/wiki/Porter_Brook/ Last accessed on 17 Jul 2013.

Yang, M. and Kang, J. (**2010**). "Psychoacoustic evaluation of various natural and artificial sounds in soundscapes." in Proceedings of European Acoustics Association (EAA) 1st EuroRegio Congress on Sound and Vibration (EAA EuroRegio 2010), Ljubliana, Slovenia, paper 135.

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

Yang, M. and Kang, J. (**2013a**). "Psychoacoustical evaluation of natural and urban sounds in soundscapes." The Journal of the Acoustical Society of America **134**(1), 840-851.

Yang, M. and Kang, J. (**2013b**). "Applicability and application of music features in soundscape." in Proceedings of AIA-DAGA 2013 International Conference on Acoustics, Merano, Italy, 1477-1480.

Yang, W. and Kang, J. (**2003**). "A cross-cultural study of soundscape in urban open public spaces." in Proceedings of the 10th International Congress on Sound and Vibration, Stockholm, Sweden, 2703-2710.

Yang, W. and Kang, J. (**2005a**). "Acoustic comfort evaluation in urban open public spaces." Applied Acoustics **66**(2), 211-229.

Yang, W. and Kang, J. (**2005b**). "Soundscape and sound preferences in urban squares: A case study in sheffield." Journal of Urban Design **10**, 61-80.

Yu, L. (**2003**). *Predicting Sound Field and Acoustic Comfort in Urban Open Spaces using Neural Networks.* MSc thesis, The University of Sheffield, Sheffield, UK.

Yu, L. and Kang, J. (**2008**). "Effects of social, demographical and behavioral factors on the sound level evaluation in urban open spaces." The Journal of the Acoustical Society of America **123**(2), 772-783.

Yu, L. and Kang, J. (**2009**). "Modeling subjective evaluation of soundscape quality in urban open spaces: An artificial neural network approach." The Journal of the Acoustical Society of America **126**(3), 1163-1174.

Zeitler, A. and Hellbrück, J. (**2001**). "Semantic attributes of environmental sounds and their correlations with psychoacoustic magnitudes." in Proceedings of the 17th International Congress on Acoustics (ICA 2001), Rome, Italy, 114.

Zwicker, E. (**1958**). "Über psychologische und methodische grundlagen der lautheit." ("On psychological and methodological bases of loudness") Acustica **8**, 237–258.

Zwicker, E. (**1965**). "Temporal effects in simultaneous masking and loudness." The Journal of the Acoustical Society of America **38**(1), 132-141.

Zwicker, E. (**1966**). "Lautstarkeberechnungsverf ahren im Vergleich." ("A comparison of methods for calculating loudness level") Acustica **17**(5), 278-284.

Zwicker, E. and Fastl, H. (**1999**). *Psychoacoustics - Facts and Models*. Springer, Berlin.

Zwicker, E., Fastl, H. and Dallmayr, C. (**1984**). "BASIC-program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532B." Acustica **55**(1), 63-67.

Zwicker, E. and Scharf, B. (**1965**). "A model of loudness summation." Psychological Review **72**(1), 3-26.

Zwicker, E., Terhardt, E. and Paulus, E. (**1979**). "Automatic speech recognition using psychoacoustic models." The Journal of the Acoustical Society of America **65**(2), 487-498.

Appendix I: Data for parameters setting of the pitch algorithms

Table AI.1 Summary autocorrelation functions (SACFs) in the pitch algorithm based on spectral method

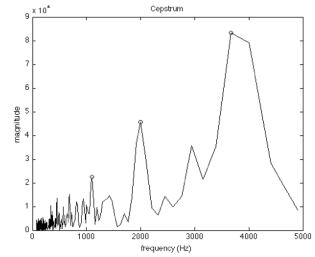
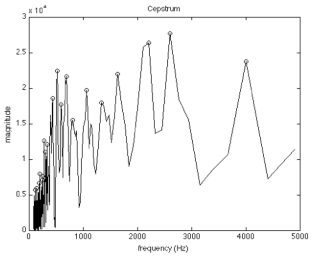
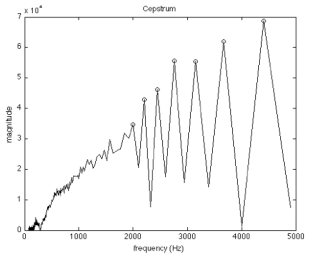
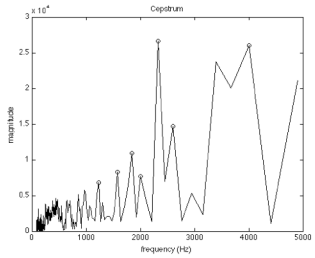
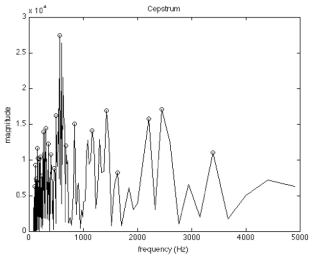
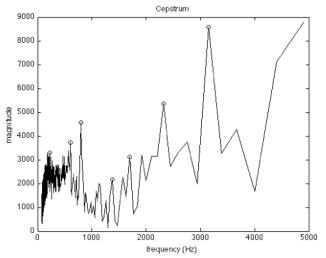
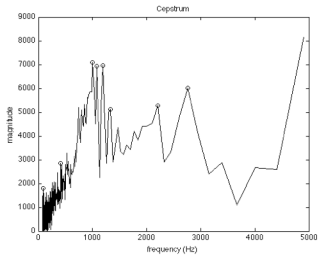
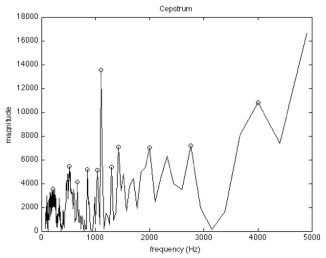
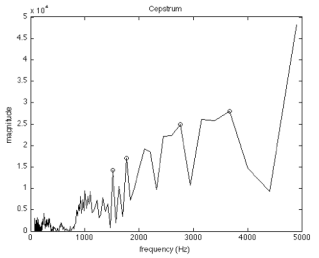
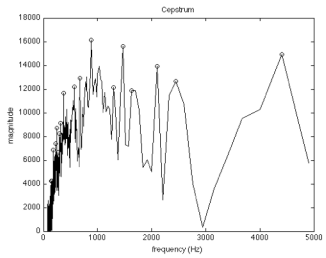
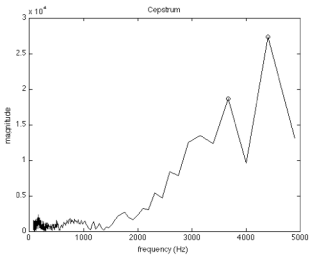
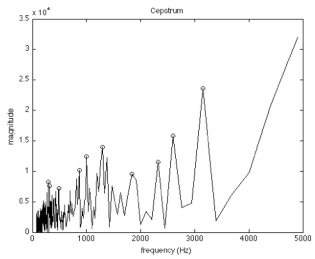
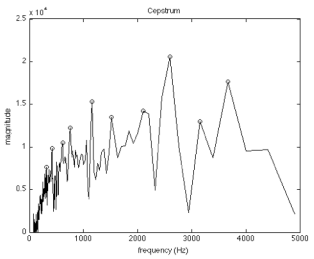
Cepstrum *[p,a]=mirpitch('folder','Cepstrum','Max',5000,'Threshold',0,'Contrast',0.2,'Reso','SemiTone')*		
		
Birdsong	Church bells	Fountain
		
Machine	Music	River
		
Sea waves	Stream	Traffic
		
Voice	Wind	
		
Big Ben	Fountain in Victoria Park	

Table AI.2 Summary autocorrelation functions (SACFs) in the modified pitch simplification algorithm

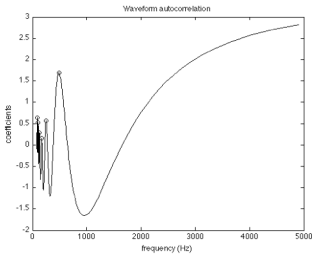
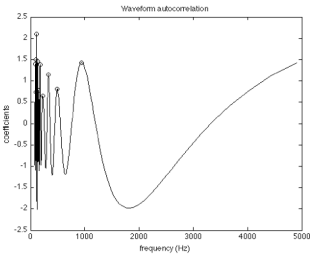
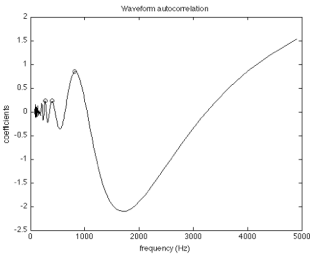
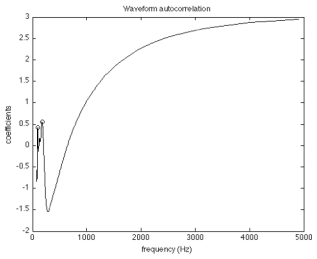
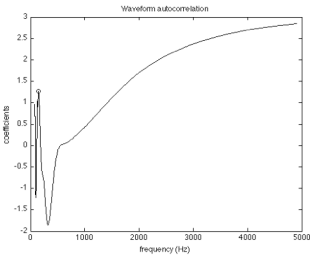
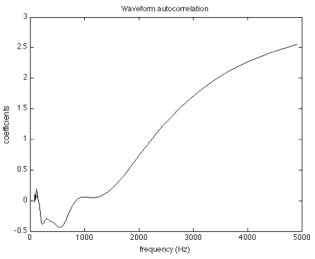
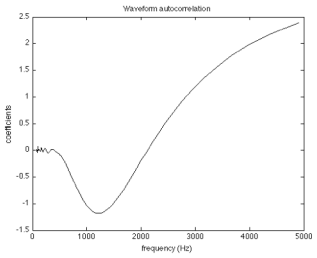
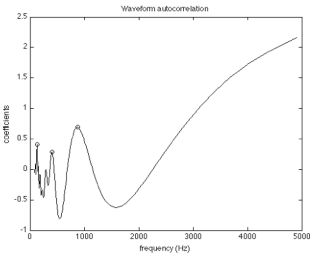
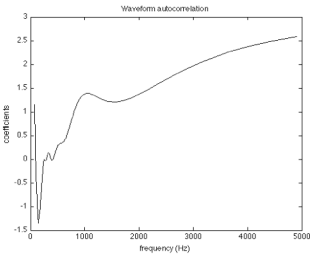
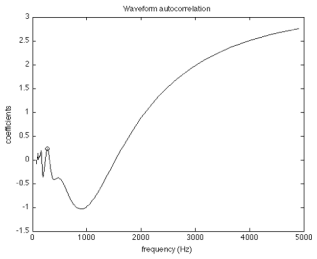
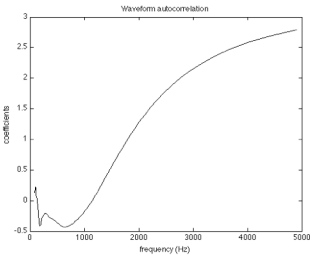
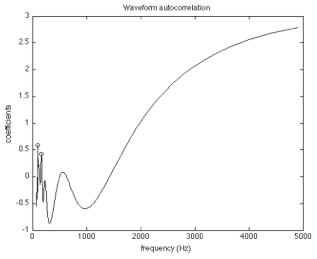
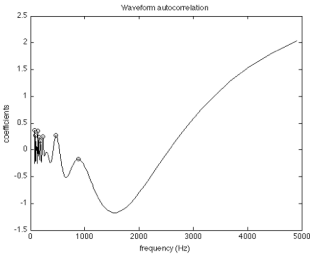
2Channels *[p,a]=mirpitch('folder','2Channels','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Generalized',2,'Enhanced',0)*		
 Birdsong	 Church bells	 Fountain
 Machine	 Music	 River
 Sea waves	 Stream	 Traffic
 Voice	 Wind	
 Big Ben	 Fountain in Victoria Park	

Table AI.3 Summary autocorrelation functions (SACFs) in the pitch algorithm based on temporal method using 10 gammatone filterbank

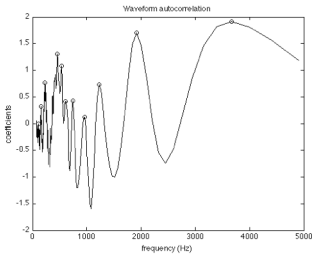
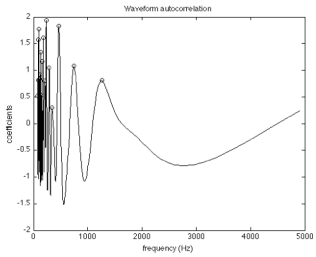
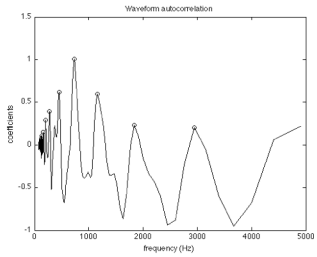
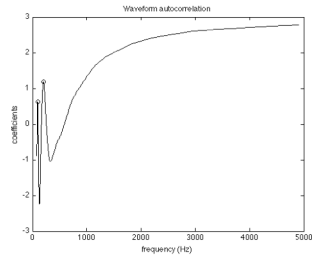
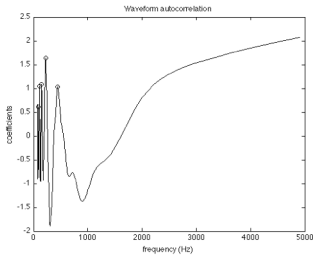
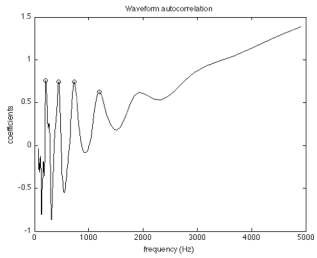
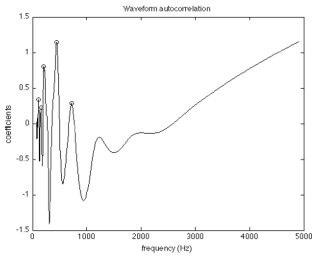
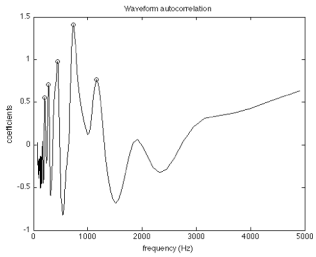
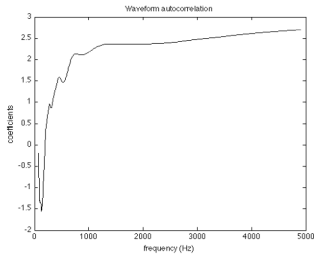
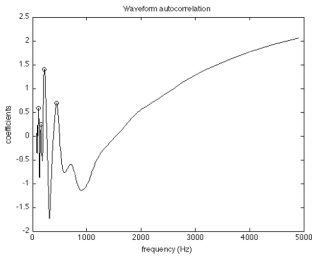
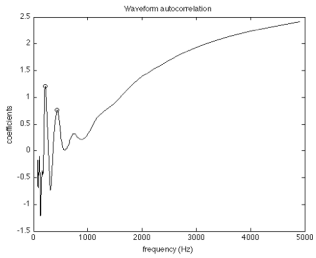
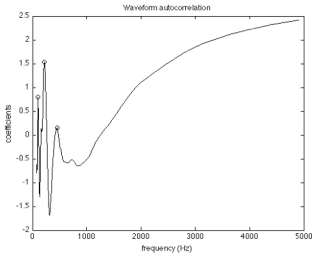
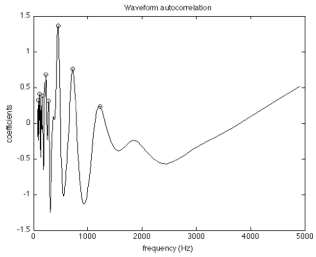
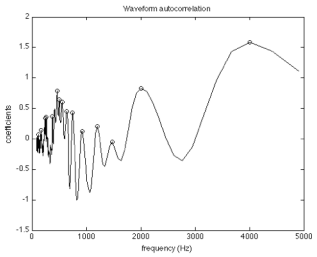
'Gammatone',10 'Hop',2 *[p,a]=mirpitch('folder','Gammatone','Max',5000,'Threshold',0.4,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0)*		
		
Birdsong	Church bells	Fountain
		
Machine	Music	River
		
Sea waves	Stream	Traffic
		
Voice	Wind	
		
Big Ben	Fountain in Victoria Park	

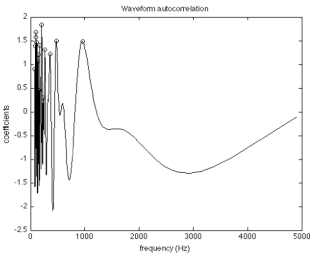
Table AI.4 Summary autocorrelation functions (SACFs) in the pitch algorithm based on temporal method using 20 gammatone filterbank

'Gammatone',20 'Hop',2

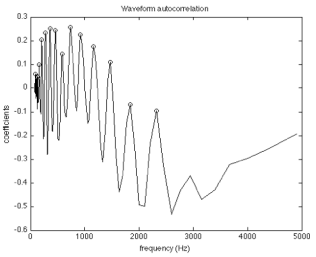
[p,a]=mirpitch('folder','Gammatone','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0)



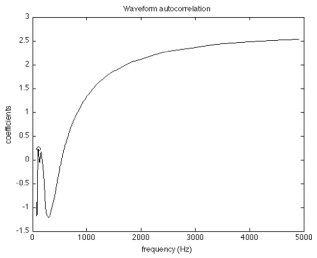
Birdsong



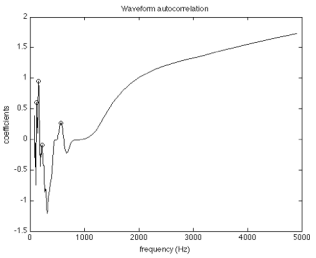
Church bells



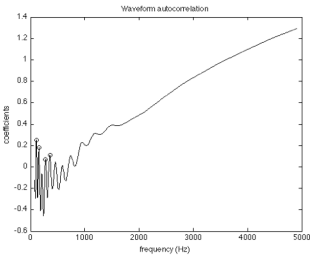
Fountain



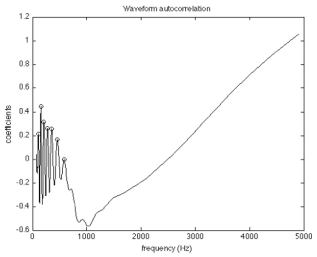
Machine



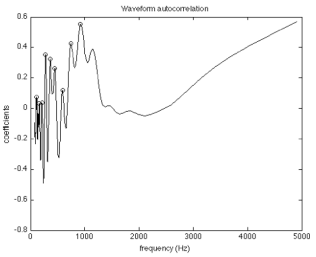
Music



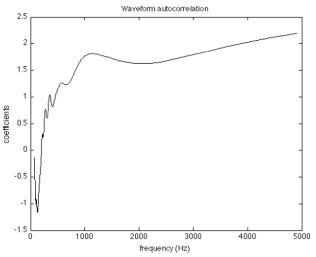
River



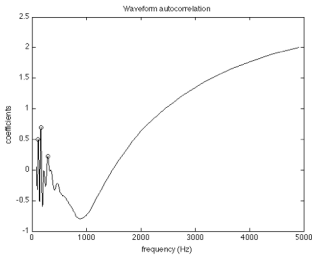
Sea waves



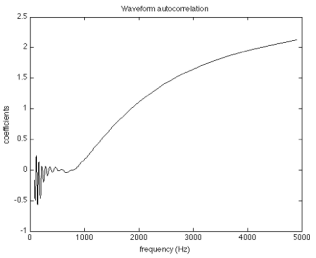
Stream



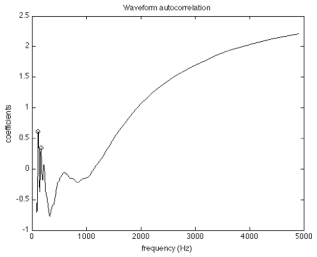
Traffic



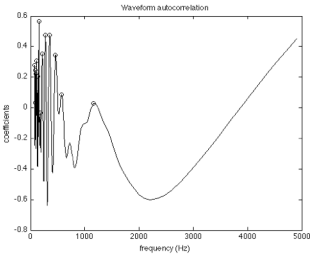
Voice



Wind



Big Ben



Fountain in Victoria Park

Table AI.5 Summary autocorrelation functions (SACFs) in the pitch algorithm based on temporal method using 40 gammatone filterbank

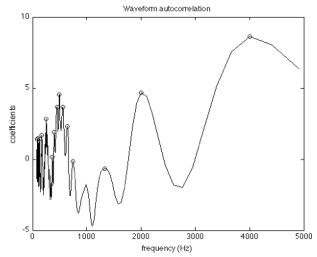
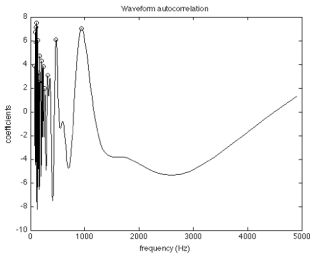
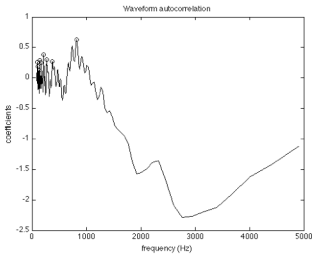
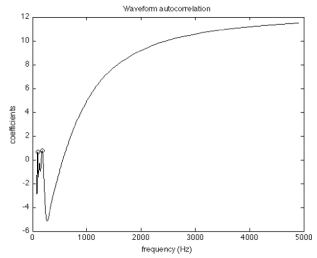
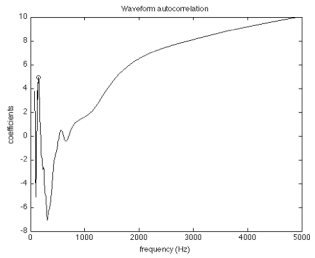
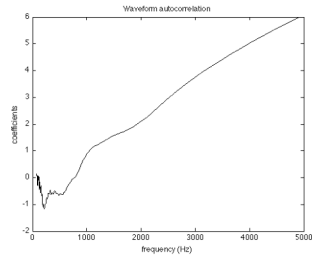
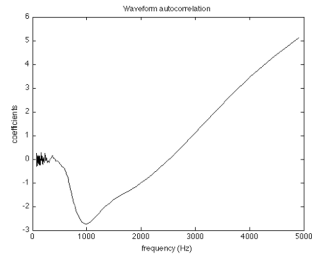
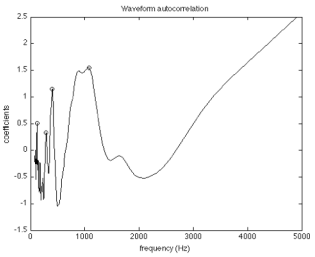
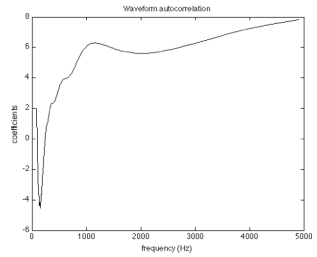
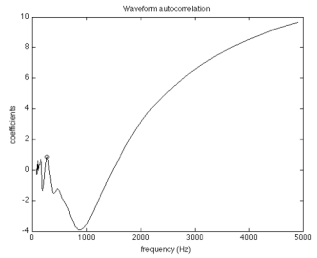
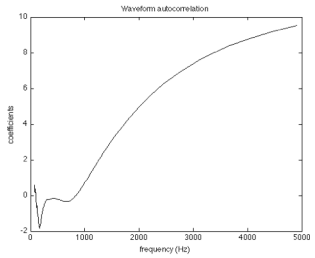
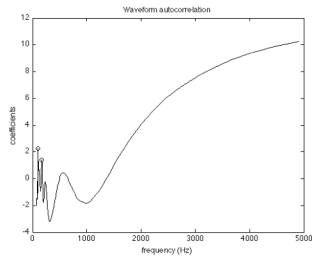
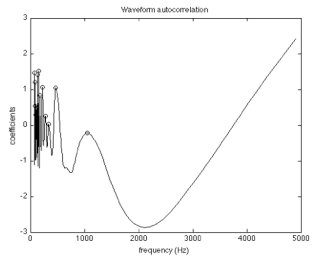
'Gammatone',40 'Hop',2 *[p,a]=mirpitch('folder','Gammatone','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0)*		
		
Birdsong	Church bells	Fountain
		
Machine	Music	River
		
Sea waves	Stream	Traffic
		
Voice	Wind	
		
Big Ben	Fountain in Victoria Park	

Table AI.6 Summary autocorrelation functions (SACFs) in the pitch algorithm based on temporal method using 80 gammatone filterbank

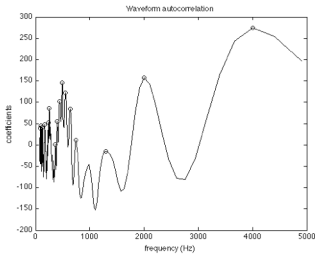
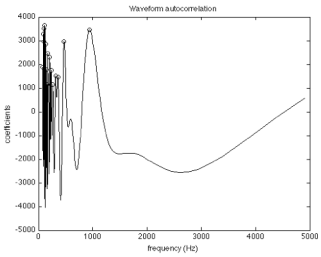
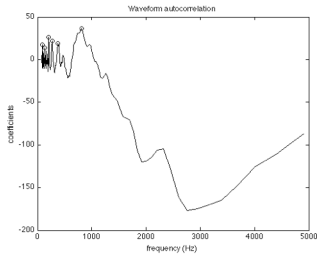
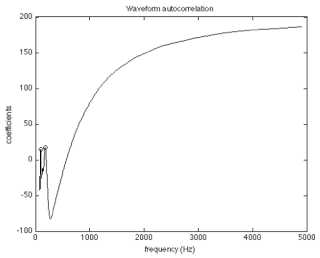
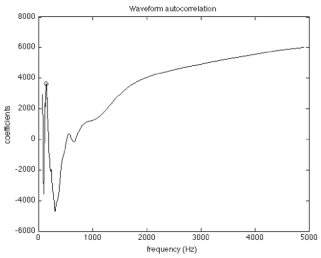
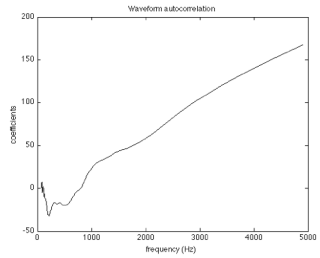
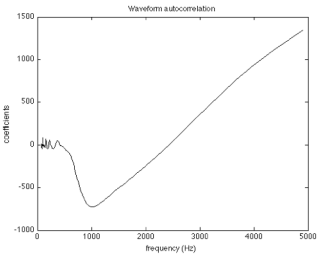
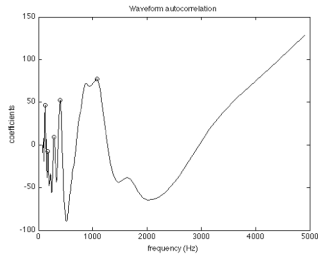
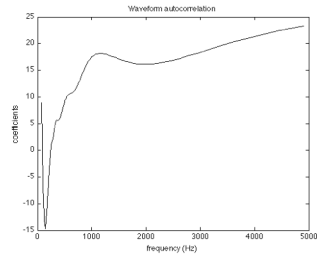
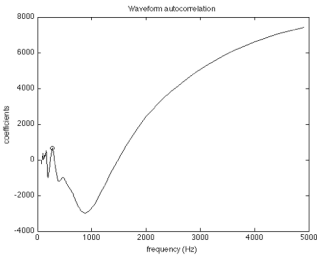
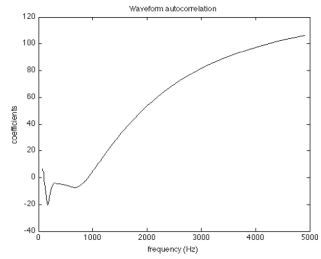
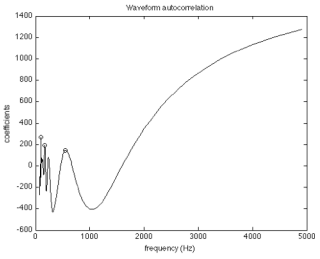
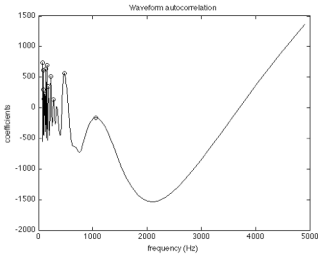
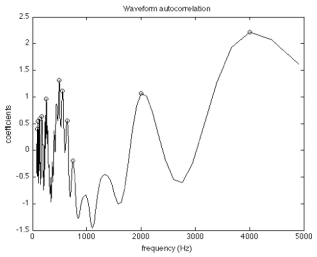
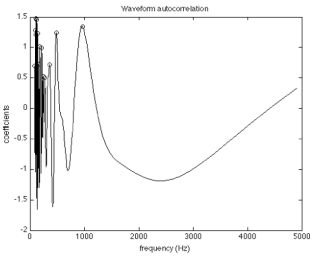
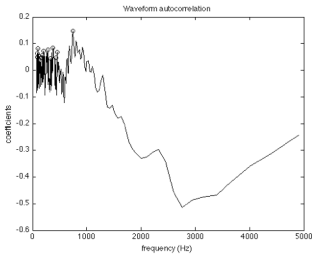
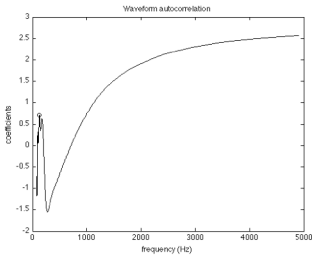
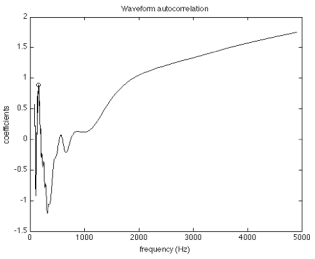
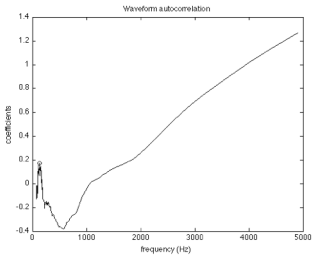
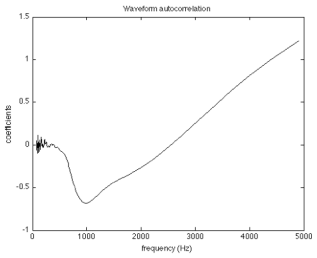
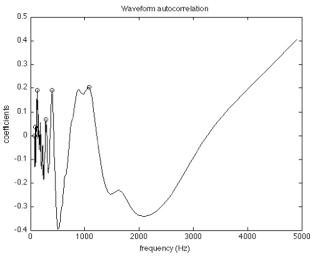
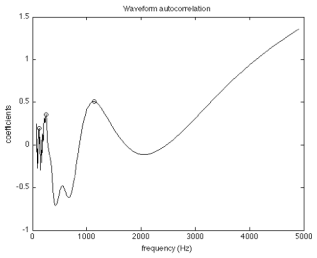
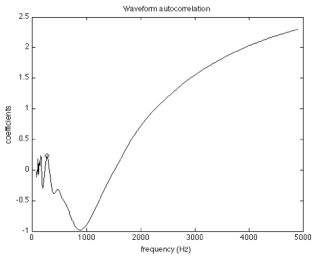
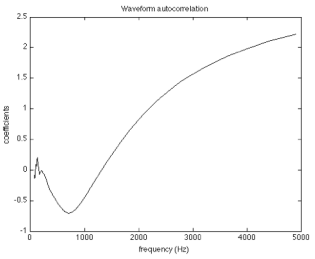
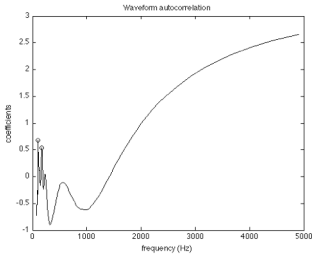
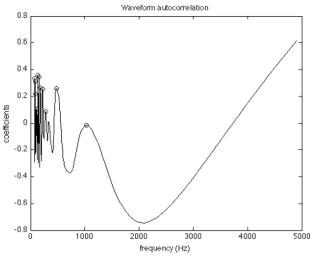
'Gammatone',80 'Hop',2 *[p,a]=mirpitch('folder','Gammatone','Max',5000,'Threshold',0.3,'Contrast',0.1,'Reso','SemiTone','Compress',2,'Enhanced',0)*		
		
Birdsong	Church bells	Fountain
		
Machine	Music	River
		
Sea waves	Stream	Traffic
		
Voice	Wind	
		
Big Ben	Fountain in Victoria Park	

Table AI.7 Summary autocorrelation functions (SACFs) in the pitch algorithm based on temporal method using Bark scale filterbanks

'Bark' *[p,a]=mirpitch('folder','Bark','Max',5000,'Threshold',0.3,'Contrast',0.15,'Reso','SemiTone','Compress',2,'Enhanced',0)*		
		
Birdsong	Church bells	Fountain
		
Machine	Music	River
		
Sea waves	Stream	Traffic
		
Voice	Wind	
		
Big Ben	Fountain in Victoria Park	

Appendix II: Data for parameters setting of the rhythm algorithms

Table AII.1 Variation of event density with time of the 13 sounds based on the envelope method of event detection

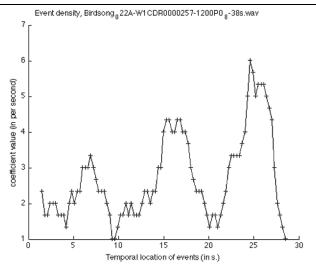
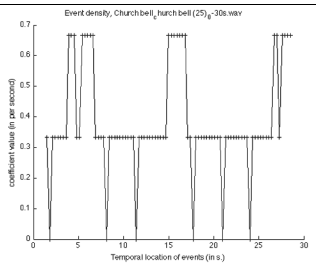
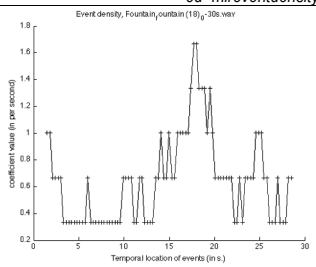
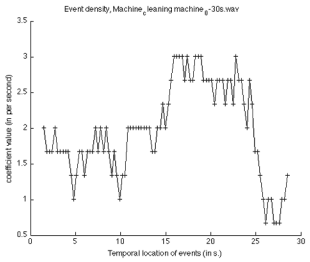
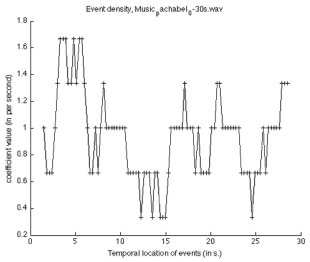
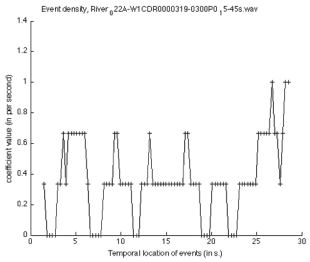
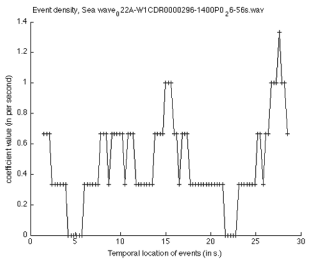
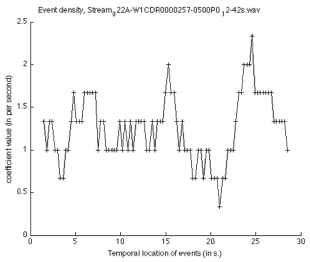
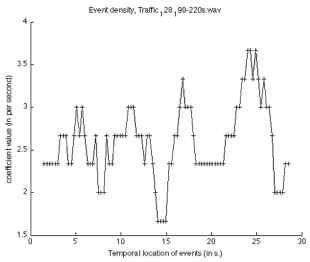
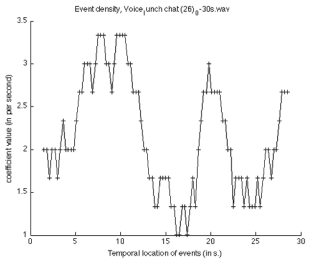
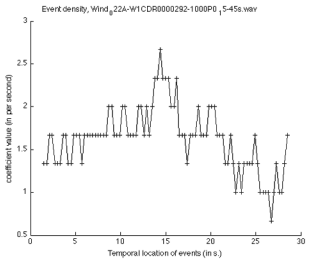
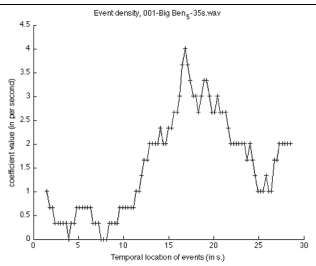
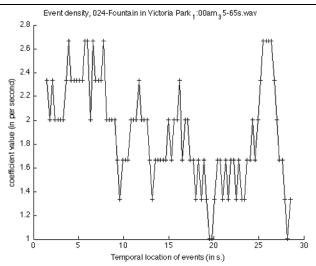
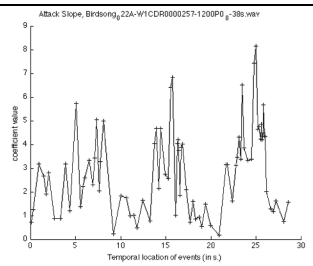
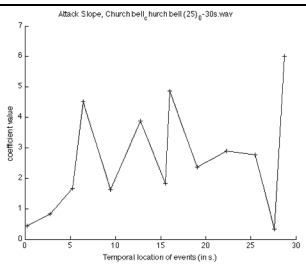
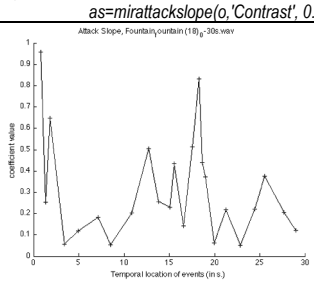
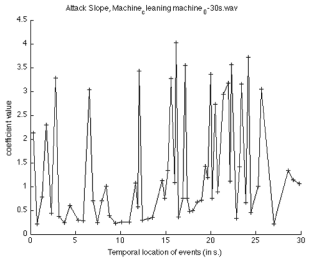
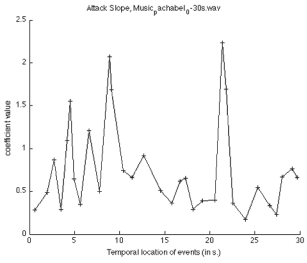
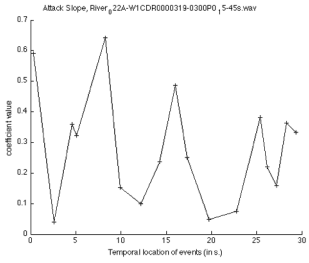
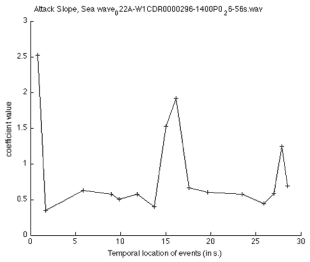
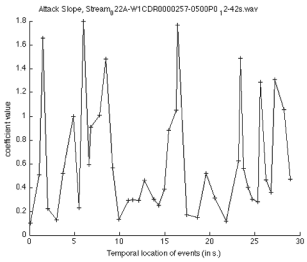
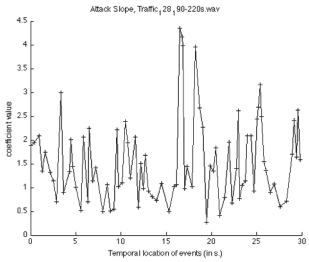
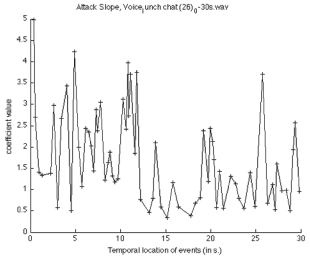
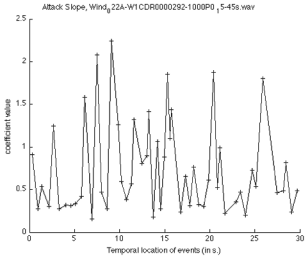
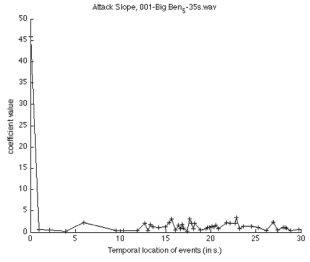
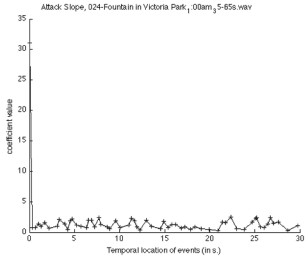
Envelope *o=mironsets('folder','Filter','Filterbank',0,'log','Contrast',0.03,'frame'); ed=mireventdensity(o)*		
		
Birdsong	Church bells	Fountain
		
Machine	Music	River
		
Sea waves	Stream	Traffic
		
Voice	Wind	
		
Big Ben	Victoria Park	

Table AII.2 Event attack slope of the 13 sounds

Envelope *o=mironsets('folder','Filter','Filterbank',0,'log','Contrast',0.03);* *as=mirattackslope(o,'Contrast', 0.03)*		
		
Birdsong	Church bells	Fountain
		
Machine	Music	River
		
Sea waves	Stream	Traffic
		
Voice	Wind	
		
Big Ben	Fountain in Victoria Park	