1-1-2005

Parametric time-frequency analysis and its applications in biomedical and multimedia

Ying Shen
Ryerson University

PARAMETRIC TIME-FREQUENCY ANALYSIS AND ITS APPLICATIONS IN BIOMEDICAL AND MULTIMEDIA

by

Ying Shen
B.A., Shanghai Institute of Foreign Trade, Shanghai, China, 1991
B.Eng., China Textile University, Shanghai, China, 1989

A thesis
presented to Ryerson University

in partial fulfillment of the
requirement for the degree of
Master of Applied Science
in the Program of
Electrical and Computer Engineering.

Toronto, Ontario, Canada, 2005

Author's Declaration

I hereby declare that I am the sole author of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Author's signature

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Author's signature

Borrowers's Page

Ryerson University requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Name	Signature	Address	Date

Abstract

Parametric Time-Frequency Analysis and its Applications in Biomedical and Multimedia

©Ying Shen 2005

Master of Applied Science
Department of Electrical and Computer Engineering
Ryerson University

Analysis of non-stationary signals is a challenging task. The purpose of this thesis is to explore an efficient and powerful technique to analyze and classify two types of non-stationary signals, that is, multimedia signals in higher frequency range (44.1 kHz) and biomedical signals in lower frequency range (2 kHz). An adaptive true non-stationary time-frequency signal analysis tool - matching pursuit, is introduced and applied to decompose the sample signals into time-frequency functions or atoms. Atom parameters are analyzed and manipulated, and discriminant features are extracted from atom parameters. Besides the parameters obtained using matching pursuit, several additional features, such as central energy and octave activeness ratio, are also derived. Linear discriminant analysis and the leave-one-out method are used to evaluate the classification accuracy rate for different feature sets. In the 6-group classification of 96 pieces of 5-second music signals, such as, christmas choir, country, greek music, jazz, rock and scottish music, the accuracy reaches 89.6%, when the feature set includes standard deviation of octave (the scale factor which controls the width of the window function), median of octave, standard deviation of innerProdI (imaginary part of the inner-product between the signal and the atom), standard deviation of realGG (real part of the inner-product between the complex atom and its conjugate), and central energy. For the database of 112 pieces of 10-second music signals, the 2-group classification (rock-like and classical-like) accuracy achieves 100%, having a standard deviation of octaves in the first 2,000 atoms as the discriminant feature. An accuracy of 74.2% is obtained for the 2-group knee sound signal classification, and optimum feature set comprises octave activeness ratio, central energy and standard deviation of innerProdI. From our experiments, it is evident that the matching pursuit algorithm with the Gabor dictionary decomposes non-stationary signals, including multimedia signals in higher frequency and biomedical signals in lower frequency ranges, into atoms whose parameters contain strong discriminant information sufficient for accurate and efficient signal classifications.

Acknowledgment

I have been exceptionally fortunate in having Dr. Sridhar Krishnan as my supervisor during my graduate study at Ryerson. Dr. Krishnan has always been supportive of my work and a strong source of inspiration and motivation. I express my profound gratitude to Dr. Krishnan for his constant help, great patience, and invaluable guidance.

I thank my thesis committee members who provided important feedback and support.

I am very grateful to Karthi Umapathy for his enormous help throughout my progress. I also would like to thank all my colleagues for their generous support, especially Serhat Erkucuk, Shahrzad Esmaili, Jiming Yang, Arunan Ramalingam, Syed Quadri, Sameh Soliman, Garabet Sinanian, Yuhong Wang, Zhenhe Chen, and Hua Yuan.

In addition, I acknowledge Emmanuel Bacry for developing the LastWave software package.

Finally, a very special thanks to the Department of Electrical and Computer Engineering, the Ontario Graduate Scholarship program and Micronet Research and Development, for the financial support.

Dedication

To my family for their love, support and encouragement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

CHAPTER 1 gives an outline of the work and the organization of the thesis. Section 1.1 and 1.2 provides a brief explanation on joint time-frequency analysis and time-frequency decomposition. Section 1.3 briefly introduces several existing time-frequency decomposition techniques. Section 1.4 depicts the proposed work and motivation. Section 1.5 introduces the adaptive decomposition tool used in the proposed work - matching pursuit. Section 1.6 illustrates the block diagram for the proposed method and describes the experiments conducted in the thesis work. The organization of the thesis is covered in the last section.

1.1 Time-Frequency Analysis

Non-stationary signals are the random signals whose statistics, such as mean and standard deviation, change over time. Since most of the real-world signals are non-stationary, the study and analysis of non-stationary signals is receiving more and more attention in the scientific community. Essentially for signal analysis, time series and frequency spectrum contain all the information about the underlying processes of signals. But by themselves, the best representations of non-stationary processes may not be well presented. Due to the time-varying behavior, techniques which give joint time-frequency (TF) information are needed to analyze non-stationary signals.

Gabor introduced the concept of atoms in his "Theory of Communication", and stated that any signal could be described as a superimposition of a large number of such atoms [1]. Atoms, also called basis functions, are signals localized in both time and frequency domains. This signal analysis method devises a joint function of time and frequency, a distribution, that will describe the energy density or intensity of a signal simultaneously in time and frequency [2]. Joint TF analysis combines time domain and frequency domain analysis to yield a more revealing picture of the temporal localization of signal's spectral characteristics. Features extracted from TF analysis contain the combined time-frequency dynamics of the given signal, as opposed to features along either the time- or the frequency-axis alone, as provided by conventional techniques [3].

One important approach in joint TF analysis is TF distribution. In TF distribution, signals are broken into their components inside a space with each component having some time, frequency and energy (or amplitude or magnitude) information attached to it. This approach best suits for non-stationary signals which need all the three axis of time, frequency and energy (or amplitude or magnitude) to represent them efficiently. The best possible distribution of the non-stationary signals is achieved by quadratic energy distributions. Among them, the Wigner distribution (WD) is the fundamental time-frequency representation, because it possesses a large number of desirable properties. But, it also exhibits nonlinear artifacts called cross-components that can interfere with the true signal auto-components. A WD cross-component appears between each pair of WD auto-components, and each WD cross-component oscillates in the joint time-frequency domain with a spatial frequency inversely proportional to the auto-component separation distance.

The inverse Fourier transform of the Wigner distribution is called ambiguity function (AF). The Fourier transform maps the WD auto-components to a region centered on the origin of the AF plane, whereas it maps the oscillatory WD cross-components away from the origin. The fact that the auto- and cross-components are spatially separated in the AF domain means that if a mask function is applied to the AF, some of the cross-components can be suppressed. The mask function is called the kernel of the TF distributions. Since there are many possible 2-d kernel functions, there exist many different TF distributions for the same signal. The class of all TF distributions obtained in this fashion is called Cohen's class [4]. Different selections of the kernels define various TF distributions in Cohen's class, such as Rihaczech, Choi-Williams and Margenau-Hill [5].

TF distributions can be only used for representation and visualization and not for modeling or analysis of the signals because these techniques are limited to represent the signals with possible optimum TF resolution instead of efficiently parameterising them [6].

Another approach of TF analysis is called TF decomposition. This approach is parametric and more suitable for modeling non-stationary signals. In the thesis work, TF decomposition is used, signals are decomposed into TF atoms, and atom parameter are analyzed and manipulated directly to extract discriminant features for signal classifications.

1.2 Time-Frequency Decomposition

TF Decomposition decomposes a signal into elementary building blocks - TF atoms, to represent the inner structure and the processes. It can better reveal the joint TF relationship and can be useful in determining the nature of the many kinds of non-stationary signals.

In such a decomposition, a signal $x[n]$ is represented as a linear combination of expansion functions $d_m[n]$,

$$x[n] = \sum \alpha d_m[n], \tag{1.1}$$

which can be expressed in matrix notation as:

$$\mathbf{x} = \mathbf{D}\alpha, \tag{1.2}$$

with $\mathbf{D} = [d_1 d_2 ... d_m ... d_M]$, where the signal $\mathbf{x}$ is a column vector $(N \times 1)$, α a is a column vector of expansion coefficients $(M \times 1)$. The set of expansion coefficients and functions in (1) provides a parametric representation of the signal.

The success of any TF modeling lies in how well it can model the signal on a TF plane with optimal TF resolution. The ideal case would be to have both time and frequency resolution as high as possible, or in other words to have infinitely small TF tile in the TF plane. However, the ideal TF resolution cannot be achieved due to the Heisenberg's uncertainty principle, which states that infinitely small widths in both time and frequency domains can not exist simultaneously. The TF tiling has to satisfy the condition:

$$\sigma_t \times \sigma_w \geq \frac{1}{2} \tag{1.3}$$

where σ_t is the time width of the TF tile, and σ_w is the frequency width of the TF tile. The TF resolution is limited to the lower bound of the uncertainty principle. It is proved that only Gabor functions or atoms (Gaussian) satisfy the lower bound condition [6].

1.3 Several Time-Frequency Decomposition Techniques

TF signal decomposition is an approach to analyze non-stationary signals by representing the signal as a sum of functions with well-defined TF properties. Different analysis techniques to decompose signals into TF atoms (or basis functions) have been developed. Fourier analysis and wavelet transform are the most common examples of such signal analysis models. However, in many cases, the basis functions are orthogonal to each other like for the cosines and sines function in Fourier and wavelets bases. Orthogonal basis functions are suitable for data compression applications, but they exhibit drawbacks for modeling non-stationary signals, for feature extraction application.

In Fourier analysis, these basis functions are cosines and sines that provide localization in frequency but not in time. Thus Fourier method provides a poor representation of time-localized signals.

Wavelet transform uses a known small wave estimating an unknown signal. Wavelets have an adaptive varying time width defined by the scaling parameter. But its "scale" is inversely tied to the "frequency", in other words, smaller scaling parameters correspond to higher frequencies, and larger scaling parameters correspond to lower frequencies. Based on Heisenberg's uncertainty principle, wavelets provides good time resolution and poor frequency resolution at higher frequencies, and poor time.resolution and good frequency resolution at lower frequencies.

The best-basis algorithm of Coifman and Wickerhauser performs function expansions over orthogonal bases from a carefully constructed dictionary [7]. The best-basis dictionary contains functions called wavelet packets for fast performance. However, the algorithm does not generalize to other dictionaries, which restricts the type of expansions that can be fulfilled. Another limitation of the algorithm is that the expansions are constrained to orthonormal bases. This restriction may prevent more accurate and efficient expansions for feature extraction. For example, the presence of strong transients within a signal can mask the presence of nearby portions of the signal with different time-frequency behaviors by causing the algorithm to choose a local basis that is well-suited to decomposing only the transients.

Vector quantization was introduced in the 1940's as an algorithm for obtaining information theoretical bounds [7]. It is now widely used in data compression. Shape-gain vector quantization is designed to approximate patterns in functions which occur over a range of different gain values. Shape-gain vector quantization is equivalent to approximating a function with the single term sum. The function is chosen from a large, highly redundant collection of basis functions called a codebook. Because of the extremely small size of the expansions, quantization algorithms employ the brute force method of trying expansions using all functions in the codebook to find the optimal one. Since the size of the codebooks needed to cover the sphere with a given density increases exponentially with the dimension of the space, the small number of terms in the expansions places a sharp limit on the dimension of the space from which functions can be approximated with an acceptable degree of accuracy. To expand large signals, such as digital audio recordings or images, the signals are first segmented into low-dimensional components, and these components are then quantized. The expansions can only represent efficiently those structures that are limited to a single low-dimensional partition. Structures that extend across the partitions require many more dictionary functions for accurate representation.

1.4 Proposed Work and Motivation

In our work, two types of non-stationary signals, i.e., music (multimedia) signals and knee sound (biomedical) signals, are being decomposed and analyzed to classify them into several pre-set categories.

A music signal often includes notes of different durations at the same time, thus even if a best local cosine basis can not well represent it. A music note may have different durations when played at different times, so a best wavelet packet basis may not be adaptive and flexible enough to represent this sound. To approximate music signals efficiently, the decomposition must have the same flexibility as the composer, who can freely choose the TF atoms (notes) that are best adapted to represent a sound [8].

Knee sound signals in the thesis work are vibroarthrographic (VAG) signals recorded during active movement of the human leg. VAG signals are non-stationary because the quality of joint surfaces coming in contact may not be the same from one angular position (or point of time) to another during articulation of the joint. VAG signals are multi-component due to the multiple sources of vibration. Further, the signal from a single source can propagate through different channels of tissue to the recording position, thus giving rise to multiple-energy components at different frequencies for a given joint angle (or time) [9].

Due to the highly non-stationary and multi-component nature of the signals, a more flexible and adaptive TF decomposition technique is needed to approximate signals and extract the features for classification. In this work, we use a true non-stationary tool, matching pursuit (MP) with Gabor dictionary, to deal with the signals whose localizations in time and frequency vary widely, and propose a parametric analysis method to study the atoms obtained from the decomposition and extract the discriminant features from the atom parameters.

1.5 Introduction to Matching Pursuit

Matching pursuit is a greedy algorithm; rather than finding a globally optimal expansion which is an NP (nondeterministic polynomial) hard problem, at each step of operation it finds an optimal one-element expansion. The residual of the one element expansion is then expanded in the next iteration, and so on.

It is already proved that only Gabor functions or atoms (Gaussian) satisfy the lower bound condition of Heisenberg's uncertainty principle. Thus, atoms in Gabor dictionary can reach the best possible TF resolution. Gabor dictionary is also more flexible and adaptive than wavelets since there is no restriction on windowing patterns and the scaling parameter is independent of frequency.

Because the pursuit is a multi-stage process, the dictionaries used for the expansions do not need to be as enormous as those used for single stage vector quantization. Matching pursuit is therefore capable of decomposing functions in very high dimensional spaces without requiring enormous computational resources or partitioning into orthogonal sub-spaces. Moreover, because high dimensional spaces do not need to be partitioned, the dictionaries can include structures which are much more delocalized than signal stage vector quantization codebooks can.

Unlike the best-basis algorithm, matching pursuit can use arbitrary dictionaries. In the proposed project, matching pursuit is used to perform adaptive TF decompositions of non-stationary signals. For this application, a collection of translated, modulated, and scaled Gaussians is used as the dictionary. This collection, optimally localized in time-frequency, is called Gabor dictionary. Gaussian functions have better time-frequency localization than wavelet packets. Since the expansions are not constrained to orthonormal bases, matching pursuit is better adapted to the time-frequency localization of signal structures. Thus using matching pursuit with Gabor dictionary is more efficient than the best-basis algorithm to perform non-stationary signal decomposition.

1.6 Applications of Matching Pursuit in the Thesis Work

In this work, three experiments, two on multimedia signal databases and one on a medical signal database, are conducted to illustrate the usefulness of matching pursuit as a signal decomposition tool for analysis and classification of non-stationary signals.

Figure 1.1 shows the schematic representation of the feature extraction, selection, and classification systems used in the thesis work.

Each non-stationary signal is decomposed into atoms using matching pursuit. Atom parameters are analyzed and manipulated to obtain discriminatory information. Discriminant features are extracted from the parameters. In order to automatically group signals of same characteristics using the discriminatory features derived, pattern classification is carried out using linear discriminant analysis (LDA) technique. The leave-one-out method is employed to estimate the correct classification rate with a least bias.

Three experiments are conducted on three databases: two of music (multimedia) samples and one of knee sound (biomedical) samples, to find the discriminant features and evaluate the classification performance.

Two different music sample databases are constructed. According to the human neurological behavior examined by Perrot et al. [10], a 5-second excerpt is long enough for people to identify the specific instrumentation, rhythm, tune, amplitude, pitch, brightness, etc., and to classify it into one of the six categories described below. However, many CD sales sites feature clips that are about 30 seconds long, and classifications for music samples longer than 5 seconds may be in need. Thus, in order to make the proposed work more practical and convincing, the second database is comprised of music samples that are 10-second in duration.

The first experiment analyzes a database of 96 music samples. The samples are obtained using single-channel recordings and the sampling rate is 44.1 kHz. Each sample has a duration of 5 seconds. The 96 pieces of music samples are classified into 6 groups, i.e. christmas choir, country music, greek music, jazz, rock, and scottish music.

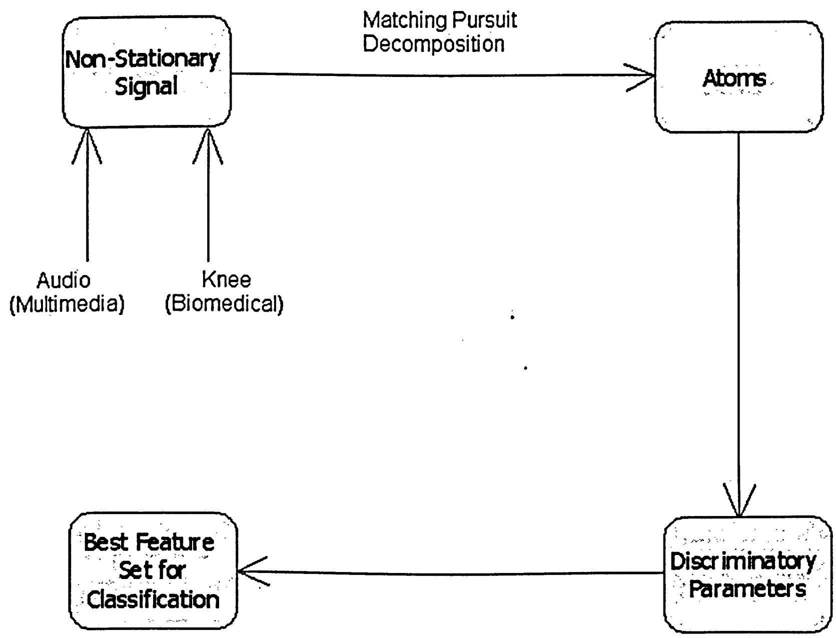

Figure 1.1: Block diagram of the proposed method for non-stationary signal classification.

In the second experiment, a database comprised of 112 pieces of music samples is decomposed and classified into two group, rock-like music and classical-like music. Each music signal has the same sampling rate of 44.1 kHz, while the duration of each sample has doubled from the first experiment, that is, 10 seconds.

The third experiment deals with a database of 89 knee sound signals and classifies them into two categories, normal and abnormal. The knee sound signals also known as Vibroarthrographic (VAG) signals are pre-filtered and has a sampling rate of 2 kHz.

From the three experiments, it is shown that the proposed method (see Figure 1.1) analyzes and classifies different ranges of non-stationary signals with acceptable accuracy. Without any signal segmentations, matching pursuit decomposes the whole non-stationary signal into atoms, and the efficient classification feature sets are found by analyzing the atom parameters. The basis function dictionary employed in all the experiments is Gabor dictionary. The proposed decomposition algorithm and the Gabor dictionary employed work well on both lower frequency signals, such as knee sound signals (at 2 kHz), and higher frequency signals, such as music samples (at 44.1 kHz). The method also works well on 2-group and multi-group classifications (6-group in the first experiment). The thesis is one of the very few works that analyze atoms statistically and extracts discriminant features directly from the parameters. This work demonstrates that the proposed method could be an efficient technique for non-stationary signal analysis and classification useful in different applications, and further study on atoms retrieved from matching pursuit and the associated parameters should be worthwhile and rewarding.

1.7 Thesis Outline

The thesis is organized into five chapters. Chapter 2 explains in detail the methodologies and techniques employed in this study. Chapter 3 demonstrates the classification experiments carried out on two music (multimedia) databases, including signal decomposition, parameter analysis, feature selection and testing, and the estimation of classification accuracies. Chapter 4 documents the application of the proposed methods on a medical database, the methodology and the results. Chapter 5 concludes the thesis with a summary of the results obtained, comparison to other works and further discussions, and the main contributions achieved.

Chapter 2

Techniques

IN this chapter, the signal decomposition algorithm matching pursuit and the statistic pattern recognition technique linear discriminant analysis (LDA), employed in this work, are introduced and studied.

2.1 Signal Decomposition Technique: Matching Pursuit with Gabor Dictionary

2.1.1 Atoms and Dictionary

In order to avoid signal segmentation and use a true non-stationary method to analyze non-stationary signals, we seek to approximate signals with linear combinations of a small number of atoms, i.e. basis time-frequency (TF) functions, selected from a dictionary, which is a collection of such functions.

Developed by Mallat and Zhang [11], matching pursuit (MP) is a powerful decomposition algorithm for non-stationary signals. MP decomposes signals into a linear expansion of atoms which are well localized both in time and frequency. Atoms are selected from a pre-defined over-complete dictionary, which can either suitably be modified or selected based on the application at hand. In the proposed project, matching pursuit with Gabor dictionary is used. Gabor functions are Gauss-modulated sines and have optimal localization in time and frequency. This dictionary is over-complete, including functions with a wide range of time-frequency localization and suitable for general decomposition purposes.

The basis TF functions in Gabor dictionary are generated by scaling, translating and modulating a single Gaussian function $g(t)$. $g(t)$ is supposed to be real, continuously differentiable and $O(\frac{1}{t^2+1})$. $\| g \|$ is imposed to be 1. The integral of $g(t)$ is non-zero and $g(0) \neq 0$. For any scale $s > 0$, frequency modulation ξ and translation u, we denote $\gamma = (s, u, \xi)$ and define

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g(\frac{t-u}{s}) e^{i\xi t}. \tag{2.1}$$

In order to demonstrate how a general collection of TF atoms would be generated from a single Gaussian function $g(t)$, the effects of scaling, translating and modulating parameters to $g(t)$ is studied and illustrated respectively from Figure 2.1 to Figure 2.4.

The single Gaussian function is plotted in Figure 2.1. When the scaling parameter is set to be 4, that is, $s = 4$, the Gaussian window becomes much wider, which is shown in Figure 2.2. How the translating parameter u shifts the single Gaussian window is illustrated in Figure 2.3. In this example, the translation is set to be 3, i.e. $u = 3$. Figure 2.4 shows when frequency modulating parameter ξ is 2, the shape of the Gaussian window looks different, with a ripple on each side.

When different combinations of scaling, frequency modulating and time translating parameters work on a single Gaussian window, numerous basis functions can be generated to form a over-complete TF atom dictionary, that is Gabor dictionary. Since a Gaussian function can be transformed into very different waveforms, as shown in Figure 2.2 to Figure 2.4, the atoms in Gabor dictionary are very flexible and adaptive, and have good time-frequency localization. It is possible to approximate a non-stationary signal with an expansion of the atoms selected from Gabor dictionary.

If $g(t)$ is even, which is generally the case, $g_\gamma(t)$ is centered at the abscissa u. Its energy is mostly concentrated in a neighborhood of u, whose size is proportional to s.

Atoms are selected one by one from the dictionary, while optimizing the signal approximations (in terms of energy) at each step.

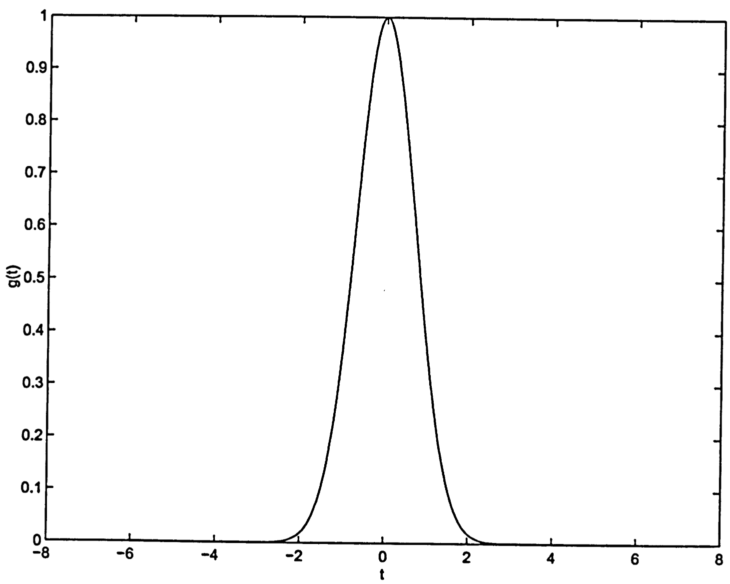

Figure 2.1: Generic plot of a Gaussian function $g(t)$.

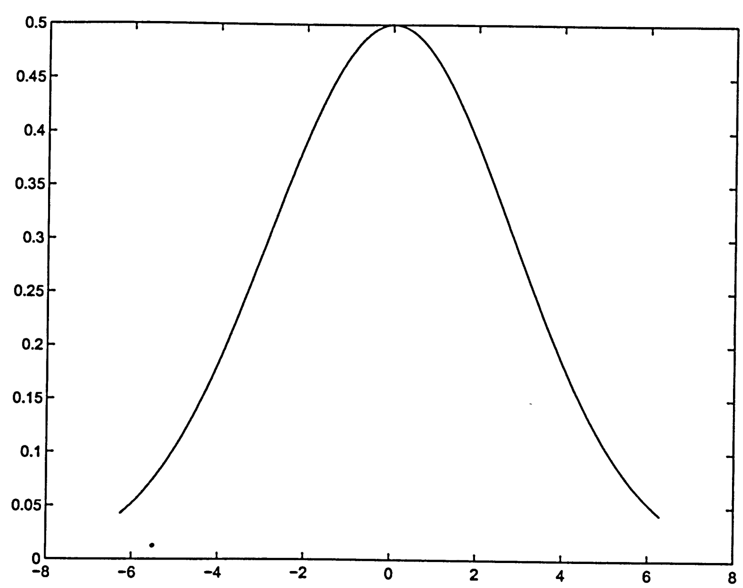

Figure 2.2: Generic plot of a scaled Gaussian function $g_\gamma(t)$, with $\gamma = (4, 0, 0)$.

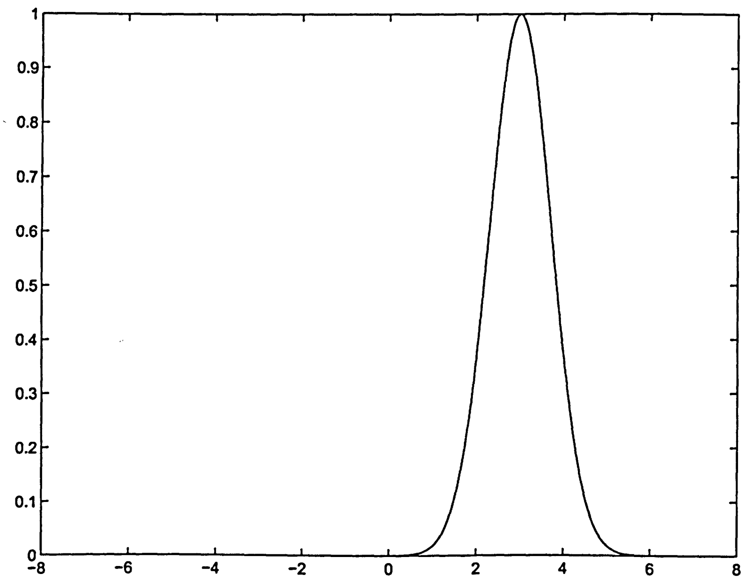

Figure 2.3: Generic plot of a translated Gaussian function $g_\gamma(t)$, with $\gamma = (1, 3, 0)$.

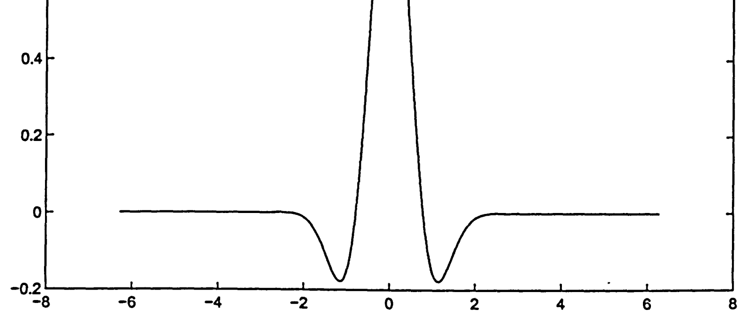

Figure 2.4: Generic plot of a modulated Gaussian function $g_\gamma(t)$, with $\gamma = (1, 0, 2)$.

2.1.2 Iterative Algorithm

Matching pursuit is a greedy signal approximation algorithm, selecting at least one atom at each iteration to best match the inner structures of a signal. At the first iteration, signal f can be decomposed into:

$$f =< f, g_{\gamma 0} > g_{\gamma 0} + Rf, \tag{2.2}$$

where $g_{\gamma 0}$ is the first atom chosen from the dictionary, Rf is the residual function after approximating f in the direction of $g_{\gamma 0}$, $< f, g >$ denotes the inner product of the signal f and the selected atom g, and $g_{\gamma 0}$ is orthogonal to Rf.

In Equation 2.2, to minimize $\| Rf \|$, $g_{\gamma 0}$ is chosen from the dictionary so that $|< f, g_{\gamma 0} >|$ is maximum. In some cases, it is only possible to find an atom $g_{\gamma 0}$ that is almost the best in the sense that

$$| < f, g_{\gamma 0} > |\geq \alpha sup \mid < f, g_{\gamma} > \mid , \tag{2.3}$$

where α is an optimality factor that satisfies $0 < \alpha \leq 1$ [12].

In the above equation, *sup* stands for 'supremum'. A value is a supremum with respect to a set if it is at least as large as any element of that set. A supremum exists in context where a maximum does not, because (say) the set is open; e.g. the set (0,1) has no maximum but 1 is a supremum. sup is a mathematical operator that maps from a set to a value that is syntactically like an element of that set, although it may not actually be a member of the set.

The choice of $g_{\gamma 0}$ is not random. It is defined by a choice function. The axiom of choice guaranties that there exists at least one choice function, but in practice, there are many ways to define it, which depends on the numerical implementation.

Matching pursuit is an iterative algorithm that sub-decomposes the residue Rf by projecting it on a basis function in the dictionary that matches Rf almost at best, as it was done for f.

After M iterations, the signal f can be decomposed in a concatenated sum,

$$f = \sum_{n=0}^{M-1} < R^n f, g_{\gamma n} > g_{\gamma n} + R^M f, \tag{2.4}$$

where $g_{\gamma n}$ is the n^{th} basis function selected from Gabor Dictionary, with scale s_n, translation u_n and frequency modulation ξ_n, and $R^M f$ is the residual after M iterations.

Thus, signal f can be expressed as a linear expansion of M basis functions selected from the dictionary, and the residue.

2.1.3 Faster Implementation of Matching Pursuit

The main disadvantage of matching pursuit is the high computational complexity required to repeatedly calculate all the inner products and search in the over-complete dictionary for the best atom. In order to lower the computational cost and accelerate the signal decomposition process, the iterative process can be stopped before the residual component will be decomposed completely, and the search for the atoms that best match the signal residue can be limited to a sub-dictionary.

There are two ways to stop the iterative process: one is to use a pre-specified limiting number M of the time-frequency atoms, and the other is to check the energy of the residue $R^M f$. A very high value of M or a zero value for the residual energy will decompose the signal completely at the expense of increased computational complexity [3]. In this work, the pursuit iterations are pre-set to M. The signal decomposition is stopped after extracting the first M time-frequency atoms. The number of iterations M is selected according to the size of samples and the complexity of classification. As long as the atoms extracted contain sufficient discriminant information to classify the sample into the pre-set categories, a smaller number of M is preferred. Unlike in the applications of signal coding or compression where detailed messages in the signals need to be kept, in signal classification, general characteristics of signals in broad sense are sufficient. Therefore, in this work, the number of iterations is relatively small, and thus the computational complexity is relatively low.

Instead of searching in a very redundant dictionary, the search for the atoms that best match the signal residues can be limited to a sub-dictionary, which can be much smaller than the original dictionary. This faster version of matching pursuit is implemented as follows: in order to further lower the computational cost and accelerate the decomposition process, the pursuits are performed only on a set of maximum atoms which correspond to the most energetic local maxima, i.e. the small areas on the spectrogram of a signal or its residue with the highest energy concentration (both in time and frequency). When no qualified atoms are left (either because they have all been selected or because after a few iterations their energy is too low), then the corresponding spectrograms are updated (using the residual) and a new set of maximum atoms are selected [13]. The algorithm performs the pursuit on this new set and so on. To use this faster decomposition, the number of maxima in the set needs to be specified. If the number is 1, then this method is exactly equivalent to the regular matching pursuit, searching the best match in the whole dictionary. The more maxima put in the set, the faster the algorithm and the less accurate the signal approximation will be.

Since in the application of signal classification, the general characteristics of signals in a broad sense is sufficient, accuracy can be traded for efficiency. Considering the size of the samples and the complexity of classifications, a relatively large maxima is selected in this work, as long as the parameters obtained are accurate enough for classification.

2.1.4 Atom Parameters

In the three experiments, matching pursuit signal decomposition is implemented using the LastWave signal processing software package [13]. Each pursuit iteration brings out at least one atom, and each atom has 17 parameters. Among the 17 parameters, 12 are associated to atom and 5 are associated to word, which corresponds to the subspaces spanned by a few atoms (in the thesis work, the number of the subspaces is set to be 1). The 5 parameters associated to word are dim, energy in word, resEnergy, coeff2 of word, status, and the 12 parameters associate to atom are octave, timeId, freqId, chirpId, innerProdR, innerProdI, phase, g2Cos2, realGG, imagGG, energy in atom, and coeff2 of atom.

Some explanations about the parameters are as follows: Parameters associated to word:

- dim: dimension of word, i.e. the number of atoms contained in each word, for now it is always “1”.

- energy in word: Always equals to “energy in atom” in this work, as the number of atoms in word is "1".

- resEnergy: residual energy in word.
- coeff2 of word: sum of the coeff2 of atoms. Always equals to "coeff2 of atom" in this work, as the number of atoms in word is "1".
- status: always "0" in this work.

Parameters associated to atom:

- octave: the scale factor which controls the width of the window function.
- timeId: related to the discrete time samples where the atom is localized.
- freqId: related to the center frequency of the atom.
- chirpId: the chirp-rate of the atom. Always "0" in this work.
- innerProdR: the real part of the inner-product between the signal and the atom.
- innerProdI: the imaginary part of the inner-product between the signal and the atom.
- phase: used for combining multiple atoms.
- g2Cos2: always "0" in this work.
- realGG: the real part of the inner-product between the complex atom and its conjugate. "0" for most of the atoms in this work.
- imagGG: the imaginary part of the inner-product between the complex atom and its conjugate. "0" for most of the atoms in this work.
- energy in atom: energy in atom. The first extracted atom contains the largest energy.
- coeff2 of atom: equals to energy in atom in this work.

2.2 Classification Scheme

2.2.1 Linear Discriminant Analysis (LDA)

Supervised classification is conducted in the three classification experiments, as all the samples have a group label or class type associated with them [14].

In this work, pattern classification is carried out using the linear discriminant analysis (LDA) technique in SPSS statistics software package [15]. Discriminant analysis is a method to statistically distinguish between two or more groups of samples. To distinguish among the groups, a set of discriminating features are selected which measure characteristics in which the groups are expected to differ. The mathematical objective of discriminant analysis is to weigh and linearly combine the discriminating features in certain way so that the groups can be as statistically distinct as possible. In other words, the LDA method tries to find one or more linear combinations of a set of discriminating features that best separate the groups of samples. These combinations are called canonical discriminant functions and have the form:

$$f = x_1 b_1 + x_2 b_2 + ... + x_{10} b_{10} + a, \tag{2.5}$$

where $x_1 ... x_{10}$ is the set of features, $b_1 ... b_{10}$ and a are the coefficients and constant, respectively, which are estimated and derived during the LDA procedure[15].

The procedure automatically chooses a first function that will separate the groups as much as possible. It then chooses a second function that is both uncorrelated with the first function and provides as much further separation as possible. The procedure continues adding functions in this way until reaching the maximum number of functions..

2.2.2 Leave-One-Out Method

In this study, the classification accuracy is estimated using the leave-one-out method which is known to provide a least bias estimate. Leave-one-out is the degenerate case of K-fold cross validation, where K is chosen as the total number of examples. In the leave-one-out method, one sample is excluded from the dataset and the classifier is trained with all the remaining samples. Then the excluded sample is used as the test data and the classification

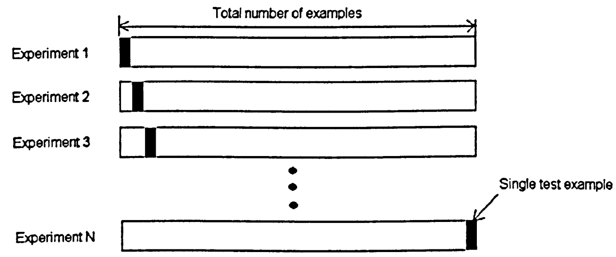

Figure 2.5: Illustration of the leave-one-out method.

accuracy is determined. Afterwords, the sample is re-entered into the dataset and a different sample is excluded and used to test the classification accuracy. This operation is repeated for all samples in the dataset. The number of correctly classified cases is used to calculate the classification accuracy rate. Since each sample is excluded from the training set in turn, the independence between the test set and the training set is maintained.

Figure 2.5 illustrates how the leave-one-out method works [16]. In a database with N examples, N experiments are performed. For each experiment, $N-1$ examples are used for training and the remaining example is used for testing. The number of correctly classified subjects is counted to estimate the classification accuracy rate. The true error is estimated as the average error rate on test examples:

$$E = \frac{1}{N}\sum_{i=1}^{N} E_i. \tag{2.6}$$

Chapter 3

Applications in Multimedia: Music Classifications

CHAPTER 3 covers the experiments on music (multimedia) signal classification. Music genres and classification applications are introduced in Section 3.1. Music database construction and processing are explained in Section 3.2. Experiments and results for 6-group music classification is given in Section 3.3, and experiments and results for 2-group music classification is covered in Section 3.4. The last section gives the conclusions and contributions.

3.1 Introduction

3.1.1 Introduction to Music Genre

Music is not only for entertainment and pleasure, but also has been used for a wide range of purposes due to its social and physiological effects. Music genres are descriptions or labels created and used by humans for classifying and describing the vast universe of music. Music genres have no strict definitions and boundaries, as they arise through a complex interaction between the public, marketing, historical and cultural factors. There are different perceptual criteria that can be used to characterize a particular music genre. Traditional music genres consist of classical, rock, jazz, country, blues, reggae, and so on. Music could be classified from pure human emotional perception, such as romantic, peaceful, sad, scary. Music could also be classified by country, composer, band, media, or other criteria.

Humans are remarkably good at genre classification, according to the human neurological behavior examined by Perrot et al. [17]. This finding suggests that humans can judge genre without any high-level theoretic descriptions. Genre hierarchies are typically created manually by human experts and are currently used to organize and structure music databases.

3.1.2 Music in Computer

The rapid increase in speed and capacity of computers and networks has allowed the inclusion of music as a data type in many applications. Music can be represented in computer in two different ways. One is based on musical scores, with one entry per note, keeping track of the pitch, duration (start time / end time), strength, and so on, for each note. Examples of this representation include MIDI (Musical Instrument Digital Interface), Humdrum [18], NIFF (Notation Interchange File Format), and Enigma [19], with MIDI being the most popular format. The other is based on acoustic signals, recording the audio intensity as a function of time, sampled at a certain frequency, and often compressed to save space. Examples of this representation include .wav, .au, and MP3 [18].

Score-based representations are much more structured and easier to handle than raw audio formats. But score-based music recordings are not as rich, expressive and beautiful as raw music ones. Most of the music data in computer is still in various raw audio formats.

3.1.3 Content-Based vs. Text-Indexed

Traditionally music databases stored in computer are organized and retrieved using one or several of the text indices, just like other textual information. Although manual indexing and classification have proved to be useful and widely accepted, finding a computerized method which allows efficient and automated classification plus easy and fast retrieval of music database is of increasing importance for several reasons:

First, manual indexing and categorization is extremely time-consuming. Multimedia databases can easily have thousands of music recordings. Radio stations and music TV channels hold archives of millions of music tapes and video clips [20]. There is also a rapid increase in the amount of music archives. It is likely that in the near future all recorded music in human history will be available on the web. The workload of manual indexing would become truly unbearable. For an easy and fast organization of databases, automatic indexing of music is required. In recent years, the interest of the research community in indexing multimedia data for retrieval purposes has grown considerably.

Secondly, because multimedia data is content in nature instead of explicit textual information, the division of music into genres is subjective and somewhat arbitrary. The quality of the division depends on the operator's knowledge, experience and personal perception. The significance of the genre relies on the mutual understanding and agreement between classifiers and listeners. In reality, misunderstanding and disagreement are unavoidable. Even different editors might have totally different opinions. Sometimes, certain classification made by one person could be totally useless to another person. Finding a content-based analysis method is the only way to create objective, accurate, and consistent classifications.

Thirdly, music genre is a language about music categorization widely used in music stores, radio stations and the Internet. In order to bridge the communication gap among musicians, editors, distributors, sellers, and consumers or users, all the parties involved need to be trained to understand the language. Text-indexed genre is only meaningful to people with the right knowledge on the meaning of the index itself. If the database were classified by content, users would be able to tell what type is their favorite and find the similar music from the database without knowing the definition of genres.

Last, music stored in computer is treated as an opaque collection of bytes with only the text labels attached, such as name, composer, year, file format, sampling rate, and so on [21]. Users trying to search and retrieve text data can be frustrated by the inability to look inside the objects. Content-based music recognition will provide a feasible solution to make the search process more transparent and efficient.

3.1.4 Possible Applications

In this thesis, a content-based music classification scheme is proposed and tested. The proposed work may have the following applications:

1. It is possible to perform the automatic music classifications and annotations.
2. It allows users to query music by style in spite of the composer. For example, the user can search for the music with both Bach and Mozart style composed by other composers.
3. It can be used in the personalized content-based music retrieval (CBMR) system based on users' preference. A CBMR system can learn users preferred music style by monitoring the users' retrieval activities and discovering the syntactic patterns from the accessed music [22].

3.1.5 Previous Works

Content-based music recognition has been receiving increasing attention in recent years. Various algorithms have been proposed. These works can be primarily separated into two classes: one deals with score-based music, and the other deals with raw music data. The former is easier to implement because score-based representations are much more structured and easier to handle, but at the same time it lacks in generalization, as only a small fraction of music data is represented in score-based formats. The later is more general and has greater significance.

Most content-based music retrieval systems operate on score-based databases. There are a few works aiming on wave-formatted music classification [23]. S. Josephson established a fuzzy expert system to distinguish timbers [24]. E. Word et al. presented a method for content-based audio files classification, in which they used duration, pitch, amplitude, brightness and bandwidth as features [21]. D. Pye applied the methods of speech processing for music classification [25]. T. Lambrou presented a study on music signal classification using statistical features in the time and wavelet transform domains, but their work was based on a small database containing only 12 musical signals, and thus the conclusion is not convincing [26].

However, most of the existing techniques do not take into consideration the non-stationary behavior of the multimedia signals while deriving the discriminating features. Samples are examined in either the time or frequency domain where it is assumed that the signals are wide sense stationary. The computational complexity for most of the existing works is relatively high. And the classifications are mostly among farther-distanced sound groups, such as speech, music and noise, or advertisement, football and news. Only a few works analyze multimedia signals in joint time-frequency domain, using true non-stationary tools to extract discriminating features, where the classifications are among different music styles which is harder than distinguishing music from other sound recordings, such as, speech or noise.

In [10], Esmaili et al. proposed a technique using short-time Fourier Transform (STFT) where features are derived directly from the time-frequency domain. 143 music signals, with 5-second duration in each signal, are classified into six genres, that is, rock, classical, folk, jazz, pop, and country. Features extracted include entropy, centroid, centroid ratio, bandwidth, silence ratio, energy ratio, and location of minimum and maximum energy. LDA is applied to test the group classification of cases. The accuracy of classification reaches 92.3% using the leave-one-out method. The proposal deals with music signals in time-frequency domain, and features extracted reflect the non-stationary properties of music signals. The computational complexity is relatively low and classification accuracy is relatively high compared to previous works. However, since short-time Fourier Transform is used in this technique, music signals are still being segmented and the determination of optimum window size brings up challenges and uncertainty in practice.

In [27], Umapathy et al. also used matching pursuit, the same adaptive time-frequency decomposition algorithm employed in the thesis work, to analyze music samples. The music samples were treated as true non-stationary signals, and no segmentations were required. No window sizes need to be determined either. A database of 64 music samples each of 5-second duration was used. In the database, there are sixteen rock-like music, sixteen classical-like music, and 32 other types of music. All the music samples were decomposed into atoms, and atom parameter octave was used to create patterns based on a similarity measure of the music signals. These patterns were used to generate templates to classify the music signals into two different categories, i.e. rock-like music and classical-like music. All the sixteen of the actual rock-like music signals were correctly classified with 100% accurate, and fourteen out of the sixteen instrumental classical-like music signals were correctly classified with a classification accuracy of 87.5%. An overall correct classification accuracy reached 90%. Some important observations were also made, such as, the octave parameter obtained as a result of time-frequency decomposition exhibits potential discriminatory ability to classify audio signals, and the octave distribution reflects the spectral similarities for the same category of signals.

3.2 Database Construction and Music Sample Processing

3.2.1 Database Construction

Evaluation of systems that try to imitate human sensory perception is difficult because there is no single well-defined correct answer [28]. And there are no standard test databases available to evaluate music classification methods. In this work, music samples are selected by the author.

Two different databases are constructed. According to the human neurological behavior examined by Perrot et al., human beings require at least 3-second excerpts to identify different musical genres with a 70% accuracy rate while the accuracy decreases to 53% for a 250ms excerpt [10]. Therefore, the first music database contains music samples with 5 seconds in duration. In our experiment, a 5-second excerpt is long enough to for people to identify the specific instrumentation, rhythm, tune, amplitude, pitch, brightness, and so on, and classify it into one of the pre-set categories, even if the melody may not be detected and grasped. However, many CD sales sites feature clips that are about 30 seconds long, and classifications for music samples longer than 5 seconds may be in need. Thus, in order to make the proposed work more practical and convincing, the second database is comprised of music samples with 10-second in duration.

The first database contains 96 music samples equally divided into six groups, that is, christmas choir, country, greek music, jazz, rock and scottish music, and each sample is 5-second in duration. The second database contains 56 rock-like and 56 classical-like music samples, and each sample is 10-second in duration. All the music samples are collected from commercial compact disks (CDs) and transformed into .wav format. CD-quality stereo recordings have two channels, each sampled at 44.1 kHz, with each sample in one channel represented as a 16-bit integer. In the experiments, only single-channel recordings are used, and sampling rate is kept as 44.1 kHz. Thus a 5-second music clip occupies 441,000 bytes and a 10-second clip occupies 882,000 bytes.

3.2.2 Music Sample Processing

Similar decomposition and analysis are conducted for all the samples in both databases to explore the efficiency and accuracy of the proposed non-stationary signal decomposition and classification method.

Since matching pursuit is a true non-stationary analysis tool, the whole music sample (5-second or 10-second) can be decomposed at one time and no signal segmentation is needed. Gabor dictionary constructed by dilating, translating, and modulating a single Gaussian window function $g(t)$ of unit norm is employed.

Since for classification purpose somewhat general characteristics of signals in a broad sense is sufficient, the fast implementation of matching pursuit is employed. The number of pursuit iterations is pre-set to control decomposition process, and local maxima is used to limit the searching area. While ensuring that the atoms extracted from each music sample are sufficient for a satisfactory classification, we try to use fewer pursuit iterations and larger local maxima, to reduce the computational complexity and achieve a better efficiency.

3.3 Experiments and Results for 6-Group Music Classifications

3.3.1 Sample Decomposition

The database is comprised of 96 pieces of music samples, each sample has the duration of 5 seconds. The samples fit into 6 categories as described in Table 3.1:

Table 3.1: 6-group music database

Group Number	Group Name	Number of Samples	Duration of Each Sample
1	christmas choir	16	5 seconds
2	country music	16	5 seconds
3	greek music	16	5 seconds
4	jazz music	16	5 seconds
5	rock music	16	5 seconds
6	scottish music	16	5 seconds

Each music sample in the database is decomposed into atoms with matching pursuit. Atoms extracted from one signal are saved in a book, which is a variable type for storing the result of matching pursuit decompositions. The number of iterations of the pursuit is set to be 1,000. Thus the book for each signal ends up with 1,000 atoms in it, except if the pursuit stops before because the residue is zero, which has not happened in the experiment. For each iteration, a set of 100 maxima is selected to accelerate the decomposition.

As discussed in Section 2.1.4, each pursuit iteration brings out one atom, and each atom retrieved has 17 parameters. The 5 parameters associated to word are dim, energy in word, resEnergy, coeff2 of word, status, and the 12 parameters associate to atom are octave, timeId, freqId, chirpId, innerProdR, innerProdI, phase, g2Cos2, realGG, imagGG, energy in atom, and coeff2 of atom.

3.3.2 Parameter Analysis and Feature Extraction

All of the 17 parameters are analyzed and plotted. It is found that some of the parameters do not carry much distinguishing information and some parameters are redundant in meaning. For instance, dim is always "1" because the number of atoms in word is always "1" in this work. Parameters of status, g2Cos2 and chirpId are always "0" in this work. Phase plots look similar to one another. Energy in word is the same in value as energy in atom, and coeff2 of word is equivalent to coeff2 of atom, as there is only one atom in each word. coeff2 of atom equals to energy in atom in this work.

In order to determine effective classification feature sets, 6 more discriminant parameters are selected and further analyzed, that is, energy in atom, octave, freqId, innerProdR , innerProdI, and realGG. These 6 parameters for the first 1,000 atoms are extracted and plotted. In the plots, the x-axis represents the number of the atoms obtained from 1,000 pursuit iterations. Atom number 1 is extracted first, containing the largest energy. The y-axis represents parameters of energy, octave, freqId, innerProdR, innerProdI and realGG, respectively. There are 16 music samples in every group. In order to see clearly the pattern distribution for each atom parameter in each music group, 16 pieces of music samples are equally separated into two subgroups, that is, the first 8 music samples and the last 8 music samples. For each music group, every parameter for the first 8 samples are depicted together to show the distribution pattern of the specific parameter, and the distribution of the same parameter for the last 8 samples are depicted together in another plot on the same page for comparison.

The six selected parameters for the first 8 samples in christmas choir group are plotted in Figure 3.1, and those for the last 8 samples in christmas choir group are plotted in Figure 3.2. The plots of the above six parameters for the first 8 samples in country music group are in Figure 3.3, and those for the last 8 samples in country music group are in Figure 3.4. The plots of the above six parameters for the first 8 samples in greek music group are in Figure 3.5, and those for the last 8 samples in greek music group are in Figure 3.6. The plots of the above six parameters for the first 8 samples in jazz music group are in Figure 3.7, and

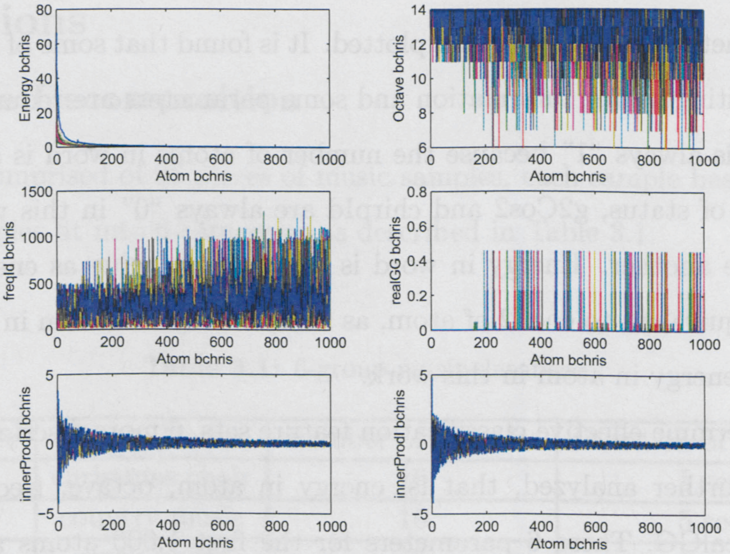

Figure 3.1: Plot of the 6 parameters for the first 8 samples in christmas choir group.

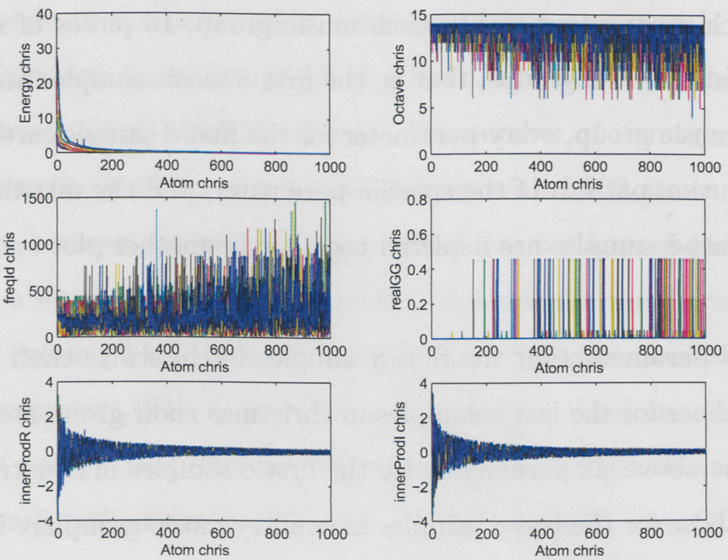

Figure 3.2: Plot of the 6 parameters for the last 8 samples in christmas choir group.

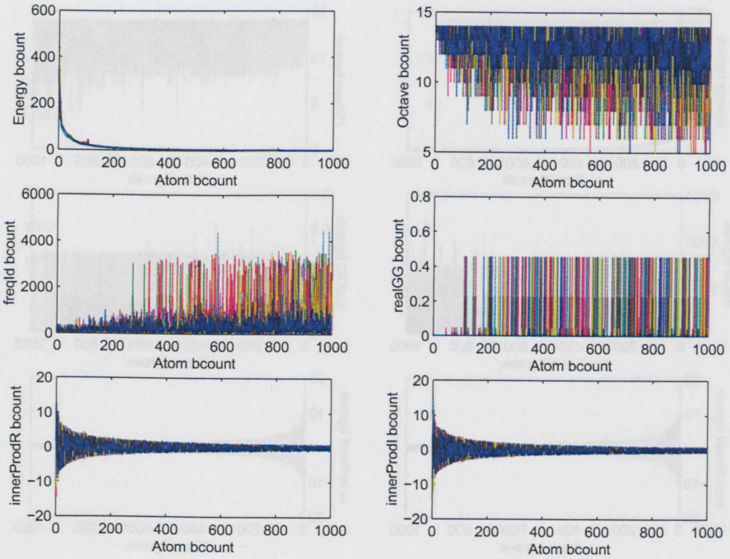

Figure 3.3: Plot of the 6 parameters for the first 8 samples in country music group.

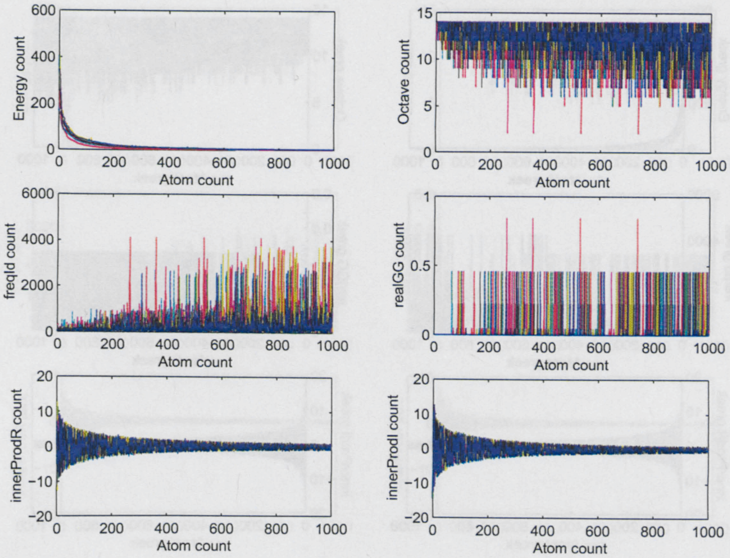

Figure 3.4: Plot of the 6 parameters for the last 8 samples in country music group.

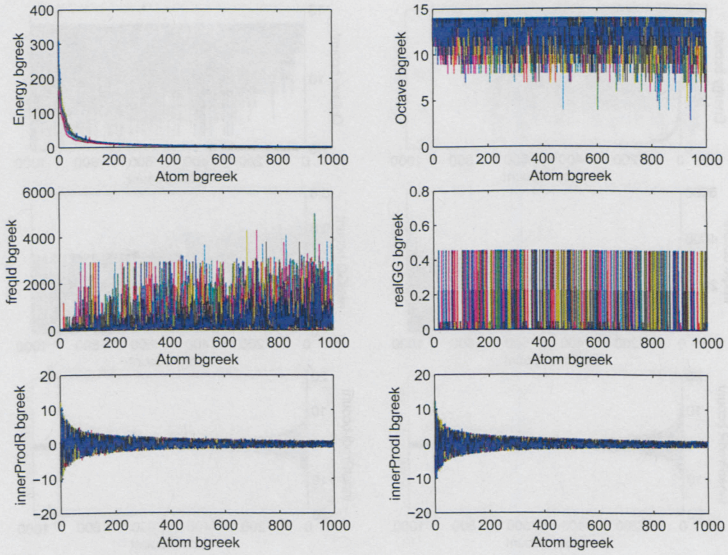

Figure 3.5: Plot of the 6 parameters for the first 8 samples in greek music group.

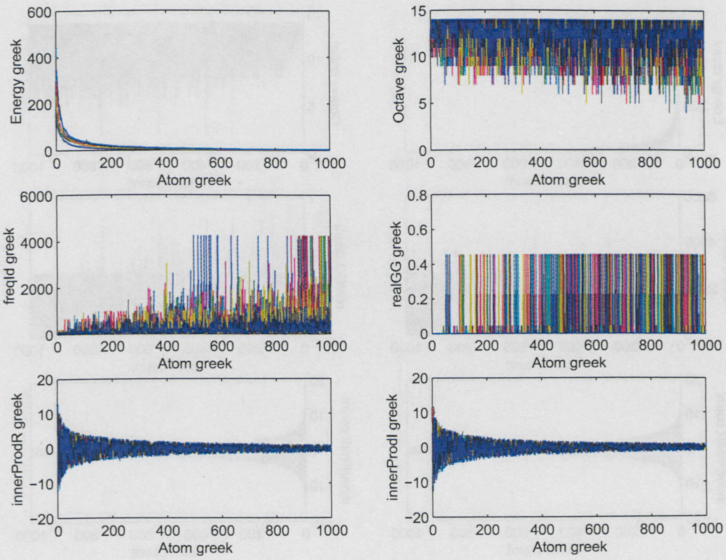

Figure 3.6: Plot of the 6 parameters for the last 8 samples in greek music group.

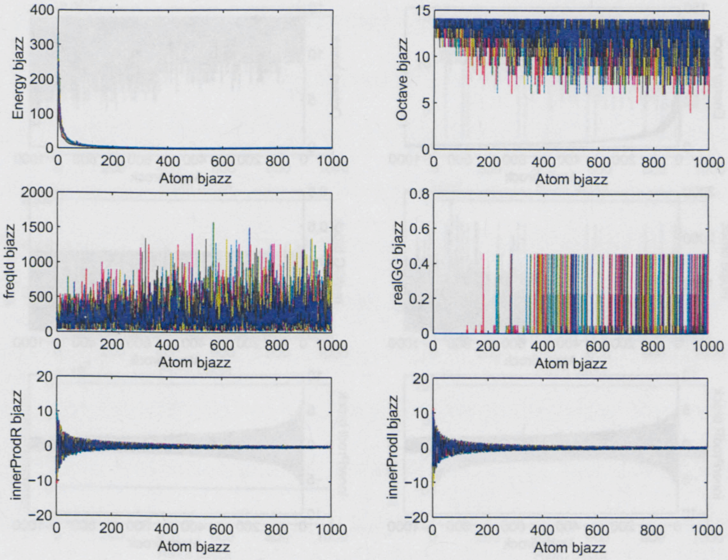

Figure 3.7: Plot of the 6 parameters for the first 8 samples in jazz music group.

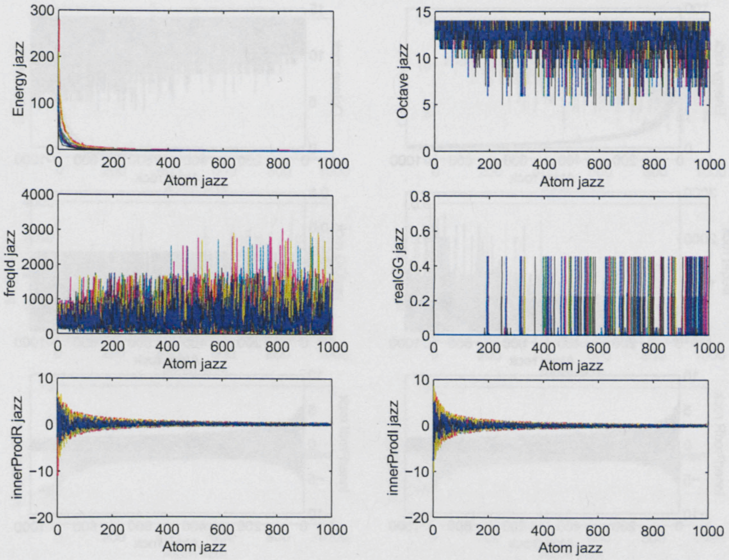

Figure 3.8: Plot of the 6 parameters for the last 8 samples in jazz music group.

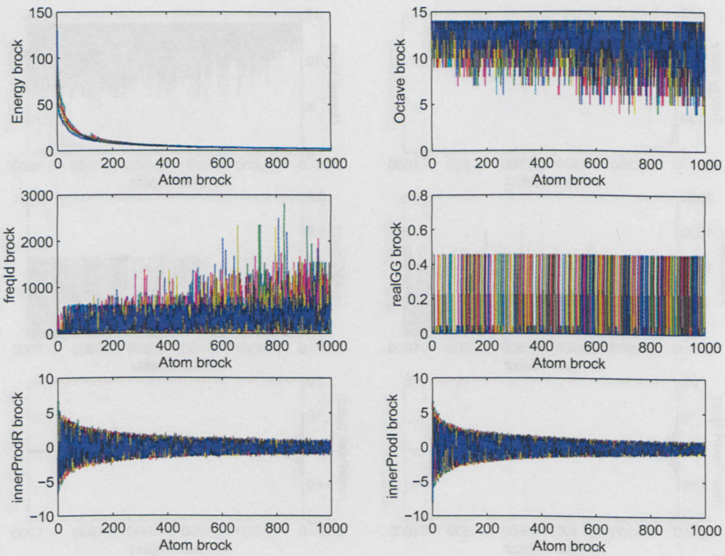

Figure 3.9: Plot of the 6 parameters for the first 8 samples in rock music group.

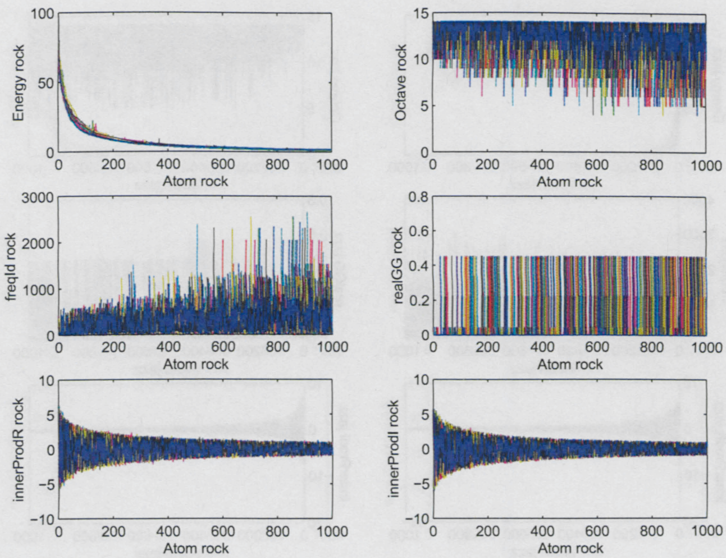

Figure 3.10: Plot of the 6 parameters for the last 8 samples in rock music group.

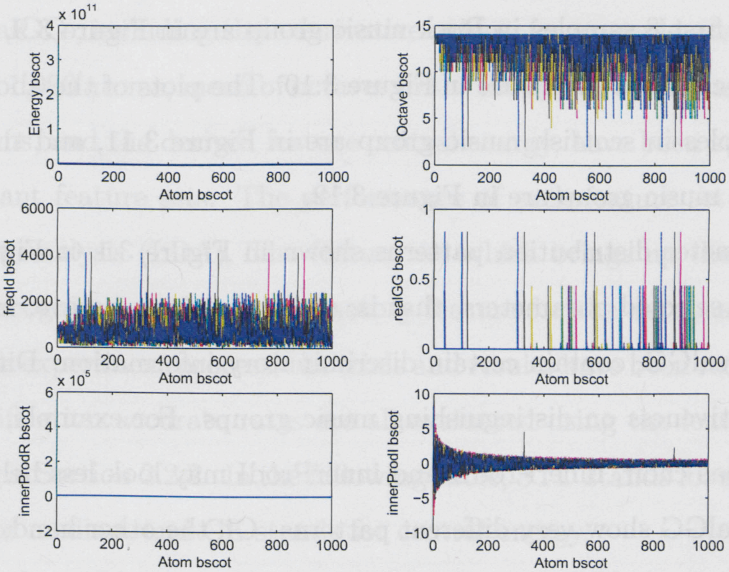

Figure 3.11: Plot of the 6 parameters for the first 8 samples in scottish music group.

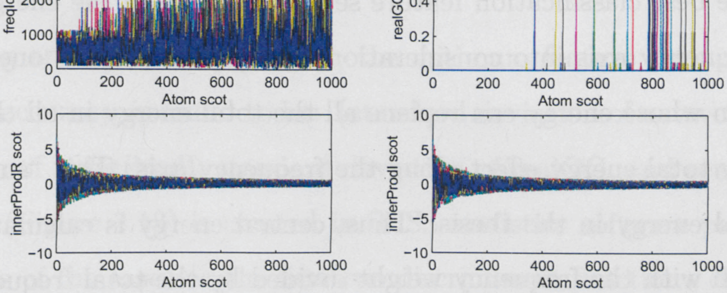

Figure 3.12: Plot of the 6 parameters for the last 8 samples in scottish music group.

those for the last 8 samples in jazz music group are in Figure 3.8. The plots of the above six parameters for the first 8 samples in Rock music group are in Figure 3.9, and those for the last 8 samples in rock music group are in Figure 3.10. The plots of the above six parameters for the first 8 samples in scottish music group are in Figure 3.11, and those for the last 8 samples in scottish music group are in Figure 3.12.

From the parameter distribution patterns shown in Figure 3.1 to Figure 3.12, it is observed that the six selected parameters, that is, energy in atom, octave, freqId, innerProdR , innerProdI, and realGG, contain certain discriminatory information. Different parameters have different effectiveness on distinguishing music groups. For example, to separate greek music from christmas choir, innerProdR and innerProdI may look less helpful, however, the distributions for realGG show very different patterns. On the other hand, to compare greek music with Rock music, realGG does not contain much distinguishing information, while the plots of innerProdR and innerProdI look quite different.

It is observed that octave and energy for the 1,000 atoms contain good discriminating information for classification. Octave is just scaling parameter and it is decided by the adaptive window duration of the Gabor function. The distribution patterns of octaves for different music groups look different. Energy distributions for the first 1,000 atoms are unique for each group. Thus, it is possible to extract good discriminant information from octave and energy in atom.

Based on the atom energy distributions, one additional feature "central energy" is derived and tried to get the best classification feature set. Having taken the energy impact in each atom along the frequency axis into consideration, we assume there is one "super" atom at a frequency location whose energy can replace all the total energy in all the atoms and still reflects the actually total energy effect along the frequency axis. This "super" atom energy is defined as central energy in the thesis. Thus, central energy is calculated as the sum of energy in each atom with the frequency weight divided by the total frequency.

In order to find an effective discriminant feature set to classify the 96 music samples into one of the six music groups, i.e. christmas choir, country, greek music, jazz, rock, and scottish music, supervised classification is conducted. The parameters of energy, octave , innerProdR, innerProdI and realGG , including their derivative values, for example, standard deviation of octaves in the first 1,000 atoms, mean of octaves in the first 1,000 atoms, median of octaves in the first 1,000 atoms, and the derived feature central energy, have been studied and selected into the discriminant feature sets. The performance of each feature set is evaluated using linear discriminant analysis (LDA). The feature set which brings up the best classification accuracy will be recognized as the discriminatory feature for the database.

To illustrate the experiment process, the trials and results have been recorded in Table 3.2. The 6-group classification accurate rates are all evaluated using the leave-one-out method, which is explained in Section 2.2.2. In the following table, STD stands for standard deviation, and Med stands for median, and CE stands for central energy.

Observation:

1) In general, combining good features can bring up better performance.

2) Sometimes, adding a good feature to the test feature set, the result is worse than adding a bad feature. For example, as an individual feature, mean of octaves provides better performance than median of octaves. However, median of octaves works better as a component in the test feature set.

3) When the result reaches a limit, adding more features to the test feature set does not necessarily bring out better results.

3.3.3 Classifications and Results

After a long trying and comparing process, the optimum feature set, which brings up the best classification accuracy, is found to be: {standard deviation of octave, median of octave, standard deviation of innerProdI, standard deviation of realGG, and central energy}.

A scatter-plot (Figure 3.13) is created in SPSS statistics software package showing the discriminant scores of the cases on the first two discriminant functions. This plot shows the separation between different cases.

All 96 music samples are categorized into six groups (christmas choir, country, greek

Table 3.2: Classification performance of the feature sets in 6-group music classification.

Feature Set	Accuracy
octave-STD	45.8%
octave-mean	51.0%
octave-Med	29.2%
innerProdI-STD	67.7%
innerProdI-mean	17.7%
innerProdI-Med	21.9%
innerProdR-STD	47.9%
realGG-STD	56.3%
CE	47.9%
innerProdI-STD realGG-STD	81.3%
innerProdI-STD innerProdR-STD	66.7%
innerProdI-STD octave-STD	72.9%
innerProdI-STD innerProdR-STD octave-STD	72.9%
innerProdI-STD octave-STD realGG-STD	87.5%
innerProdI-STD octave-STD realGG-STD CE	88.5%
innerProdI-STD octave-STD octave-mean realGG-STD CE	86.5%
innerProdI-STD octave-STD octave-Med CE	82.3%
innerProdI-STD octave-STD octave-Med realGG-STD CE	89.6%
innerProdI-STD innerProdI-mean octave-STD octave-Med realGG-STD CE	88.5%
innerProdI-STD innerProdI-Med octave-STD octave-Med realGG-STD CE	89.6%

music, jazz, rock and scottish music), and the confusion matrix depicted in Table 3.3 shows the classification performance of the optimum feature set.

All 16 pieces of christmas choir samples, jazz samples and rock samples are correctly classified. 12 out of 16 pieces of country music samples, 11 out of 16 pieces of greek music samples, and 15 out of 16 pieces of scottish music samples are well classified.

2 pieces of country music samples are mis-classified into greek music group, and another 2 pieces of country music samples are mis-classified into rock music group. 4 pieces of greek music samples are mis-classified into country music group, and 1 piece- of greek music samples are mis-classified into rock music group. 1 piece of scottish music sample is mis-classified into rock music group.

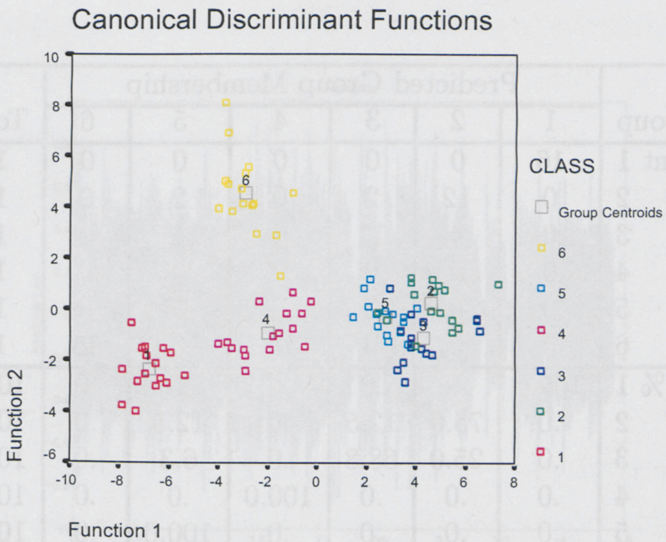

Figure 3.13: All-group scatter-plot with the first two canonical discriminant functions.

Using the leave-one-out method, 89.6% of all original grouped cases are correctly classified.

3.4 Experiments and Results for 2-Group Music Classifications

3.4.1 Sample Decomposition

The second database is comprised of 112 pieces of music samples, and each sample has the duration of 10 seconds. The samples fall into two categories, that is, rock-like music group (7 sub-groups, 8 pieces of 10-second clips in each subgroup), and classical-like music group (7 sub-groups, 8 pieces of 10-second clips in each subgroup).

Each music sample in the database is decomposed with matching pursuit. Atoms extracted from one signal are saved in a book, which is a variable type for storing the result of matching pursuit decompositions. Since each sample is doubled in size from the previous

Table 3.3: Performance of the optimum feature set in LDA classifier with the leave-one-out method.

Group	Predicted Group Membership 1	2	3	4	5	6	Total
Count 1	16	0	0	0	0	0	16
2	0	12	2	0	2	0	16
3	0	4	11	0	1	0	16
4	0	0	0	16	0	0	16
5	0	0	0	0	16	0	16
6	0	0	0	1	0	15	16
% 1	100.0	.0	.0	.0	.0	.0	100.0
2	.0	75.0	12.5	.0	12.5	.0	100.0
3	.0	25.0	68.8	.0	6.3	.0	100.0
4	.0	.0	.0	100.0	.0	.0	100.0
5	.0	.0	.0	.0	100.0	.0	100.0
6	.0	.0	.0	6.3	.0	93.8	100.0

experiment. The number of iterations of the pursuit is increased to be 3,000 to get more detailed information for effective classifications. Thus the book for each signal ends up with 3,000 atoms in it, except if the pursuit stops before because the residue is zero, which has not happened in the experiment. In order to accelerate the decomposition, a set of 300 maxima is selected for each iteration.

We try to use as few as possible atoms to reduce the computational complexity, as long as satisfying classification results can be obtained. In this experiment, the first 2,000 atoms are analyzed to find the optimum classification feature set.

Each music sample is studied. It has been observed from the sample plots that rock-like music samples tend to have larger amplitude than classical-like samples, which is in line with their acoustic features, that is, in general rock music is louder than classical music. To illustrate this point, one rock-like music sample and one classical-like music sample are randomly selected from the database and plotted in Figure 3.14 and 3.15.

In order to look more into the characteristics demonstrated by rock-like music samples and classical-like music samples, and define the discriminatory features for classification, the

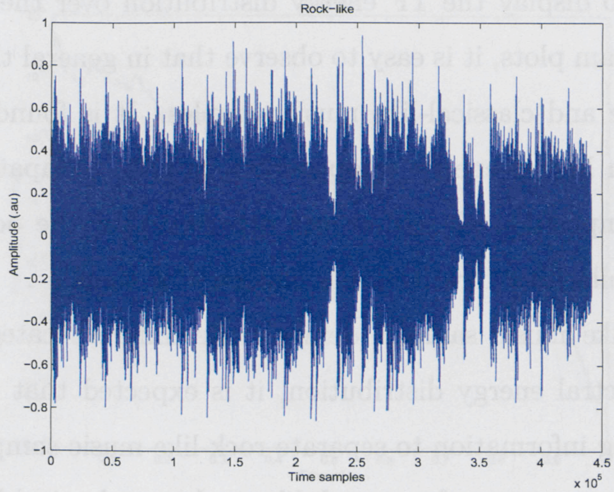

Figure 3.14: An example of 10-second rock-like music signals. *au* - arbitrary unit. Sampling rate=44.1 kHz.

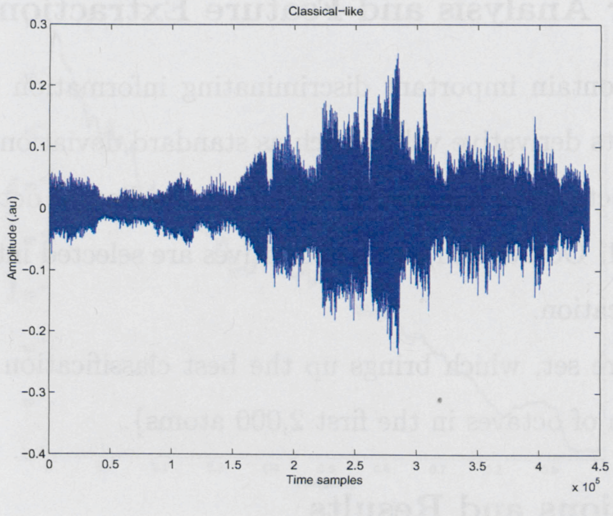

Figure 3.15: An example of 10-second classical-like music signals. *au* - arbitrary unit. Sampling rate=44.1 kHz.

spectra and spectrograms of the samples are also studied. Spectrum plots show the power spectral density. Spectrogram is the squared modulus of the short-time Fourier transform and is generally used to display the TF energy distribution over the TF plane. From the spectrum and spectrogram plots, it is easy to observe that in general the energy distribution is different for rock-like and classical-like music samples. It is found that rock-like music samples usually contain higher energy components. In [27], Umapathy et al. studied the matching pursuit decomposition algorithm and observed that the octave distribution can reflect the spectral similarities for the same category of signals . Since rock-like music samples and classical-like music samples demonstrate different categorical characteristics with regard to the spectral energy distribution, it is expected that the octave parameter may carry distinguishing information to separate rock-like music samples from classical-like ones. Spectra and Spectrograms of one rock-like and one classical-like music sample are randomly selected from the database and plotted from Figure 3.16 to Figure 3.19, to show the visible differences of the spectral energy distribution between the two groups.

3.4.2 Parameter Analysis and Feature Extraction

Knowing octave may contain important discriminating information for classification, this parameter, along with its derivative values such as standard deviation of octaves in the first 2,000 atoms, mean of octaves in the first 2,000 atoms, median of octaves in the first 2,000 atoms, has been studied. Octave and/or its derivatives are selected into the test feature sets for music group classification.

The optimum feature set, which brings up the best classification accuracy, is found to be: {standard deviation of octaves in the first 2,000 atoms}.

3.4.3 Classifications and Results

The values of standard deviation of octaves in the first 2,000 atoms are listed in Table 3.4. By observation, the threshold of 1.7 is assigned, which can separate the rock-like music samples completely from the classical-like music samples. When standard deviation of octaves in the

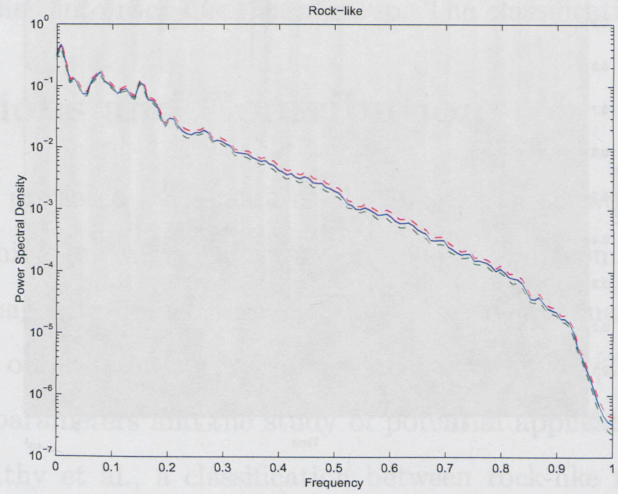

Figure 3.16: Spectrum of the rock-like music signal in Figure 3.14. X-axis: normalized frequency, maximum frequency = sampling frequency/2.

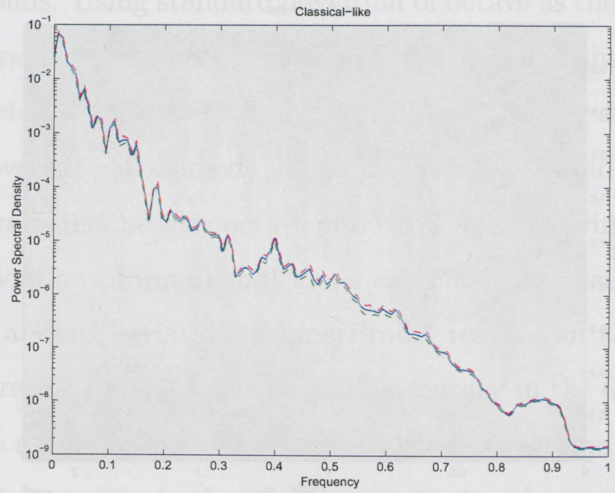

Figure 3.17: Spectrum of the classical-like music signal in Figure 3.15. X-axis: normalized frequency, maximum frequency = sampling frequency/2.

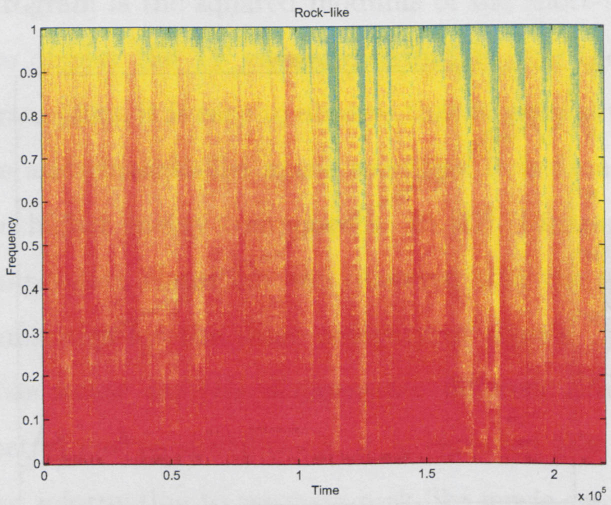

Figure 3.18: Spectrogram of the rock-like music signal in Figure 3.14. X-axis: time samples. Y-axis: normalized frequency, maximum frequency = sampling frequency/2. Colors indicate different energy levels, with blue the lowest and red the highest.

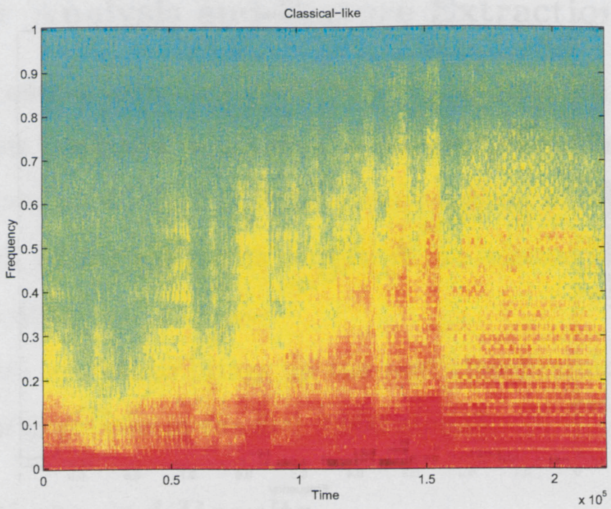

Figure 3.19: Spectrogram of the classical-like music signal in Figure 3.15. X-axis: time samples. Y-axis: normalized frequency, maximum frequency = sampling frequency/2. Colors indicate different energy levels, with blue the lowest and red the highest.

first 2,000 atoms is smaller than 1.7, the music sample is classified into classical-like music Group. When standard deviation of octaves in the first 2,000 atoms is larger than 1.7, the music sample is classified into rock-like music group. The classification accuracy is 100%.

3.5 Conclusions and Contributions

Unlike the techniques proposed by Esmaili et al. [10], the thesis work employs a true non-stationary tool, matching pursuit with Gabor dictionary, to decompose music samples, so there is no need for signal segmentation or widow size determination. The thesis work adopts some valuable observations provided by Umapathy et al. [27], and further extends the analysis of atom parameters and the study of potential applications. Compared to the work done by Umapathy et al., a classification between rock-like music and classical-like music is performed again, and the atom parameter octave is still used as the key discriminant feature. However, the database size in this work has increased, including 56 pieces of rock-like music samples and 56 pieces of classical-like music samples, and each sample has the duration of 10 seconds. Using standard deviation of octave as the discriminatory feature, the classification accuracy reaches 100%. In the first database of the thesis work, a 6-group classification is conducted. The classification performance of other atom parameters besides octave has been discovered and studied. It is observed that standard deviation values of several other atom parameters besides octave also carry good discriminant information. For example, standard deviation of innerProdI alone can classify the six music genres with an accuracy of 67.7%. Standard deviation of innerProdR reaches an accuracy of 47.9%. And standard deviation of realGG alone achieves 56.3% accuracy in the 6-group music classification. Central energy is a new feature derived in the thesis work, and it brings an acceptable accuracy of 47.9% in 6-group music classification. It is also observed that a combination of good discriminatory features may bring up improved results. For example, using standard deviation of innerProdI and standard deviation of real GG together, the 6-group classification accuracy can be increased to 81.3%. The optimum feature set, which brings up the best classification accuracy, is found to be: {standard deviation of octave, median of octave, standard deviation of innerProdI, the standard deviation of realGG, central energy}. The overall accuracy reaches 89.6%. It is also noted that adding more discriminatory features does not necessary improve the classification performance. When the accuracy rate reaches certain value, adding more features to improve the performance may become difficult. After the classification performance reaches the optimal, adding more features into the feature set may just reduce the accuracy rate.

The experiments on the music databases verify again that matching pursuit, as an adaptive time-frequency tool, decomposes non-stationary signals into atoms whose parameters contain good discriminant information for classification. The thesis work further proves that octave has the discriminatory ability to classify audio signals. It is also discovered that some other atom parameters besides octave carry satisfying discriminatory information as well. The derivative values of these parameters may act as good discriminant features, bringing good classification results. New feature central energy has good performance as well. The experiments and results may bring people's attention to more atom parameters besides octave, and inspire more interest in music classification using matching pursuit.

Table 3.4: Standard deviation of octaves in the first 2,000 atoms.

Music Sample	Standard Deviation of Octaves			
Classical 1-4	1.2109	1.1631	1.2701	1.4257
Classical 5-8	1.5357	1.4144	1.0916	1.2308
Classical 9-12	1.0760	1.2239	1.4580	1.1023
Classical 13-16	1.2622	1.1759	1.4090	1.5346
Classical 17-20	1.4979	1.4900	1.4958	1.5222
Classical 21-24	1.4492	1.6053	1.4742	1.3996
Classical 25-28	1.3389	1.2897	1.2771	1.2380
Classical 29-32	1.2351	1.2903	1.3520	1.3613
Classical 33-36	1.3665	1.2858	1.2777	1.1167
Classical 37-40	1.3031	1.4725	1.2384	1.1055
Classical 41-44	1.1702	1.1286	1.1718	1.1266
Classical 45-48	1.3096	1.1946	1.4924	1.1853
Classical 49-52	1.2886	1.1800	1.2341	1.1556
Classical 53-56	1.1894	1.2725	1.3664	1.3428
Rock 1-4	2.1355	2.3155	2.1863	2.0359
Rock 5-8	2.0105	1.9743	2.0570	2.2351
Rock 9-12	2.5278	2.5570	2.3779	2.1647
Rock 13-16	2.2028	2.2540	2.1758	2.0557
Rock 17-20	1.9922	2.0358	2.0630	1.7830
Rock 21-24	2.0853	1.9753	2.0233	1.9941
Rock 25-28	2.0534	1.9518	1.9035	1.9630
Rock 29-32	2.0667	1.8370	1.8492	1.8096
Rock 33-36	2.1048	1.9141	1.8272	1.7141
Rock 37-40	2.0565	2.0237	1.9021	1.7591
Rock 41-44	2.7277	2.5827	2.3621	2.6165
Rock 45-48	2.5539	2.6482	2.6736	2.3581
Rock 49-52	2.4693	2.3978	2.2018	2.1915
Rock 53-56	2.2678	2.1218	2.0843	2.1882

Chapter 4

Applications in Biomedical: Knee Sound Classifications

CHAPTER 4 covers the experiment on knee sound (biomedical) signal classification. Knee injuries and vibroarthrographic (VAG) signals are introduced in Section 4.1. The process of sample decomposition and feature selection is given in Section 4.2. The classification and results are covered in Section 4.3. The last section gives the conclusions and contributions.

4.1 Introduction

4.1.1 Knee Injuries and Detection

The knee joint is the largest of human joints in terms of the area of articular cartilage and synovial membrane, and the most complicated joint in terms of internal movement and mechanics. Knee joint injuries are the most common injuries to the human body as the knee joint is exposed to severe angular and torsional strains, especially in athletics. The number of surgeries performed on the knee joint is far more than the sum of all surgeries performed on other joints in the body.

The knee joint has four distinguishing features - a joint cavity, articular cartilage, a synovial membrane, and a fibrous capsule [29]. Articular cartilage serves as a wear-resistant, nearly frictionless, load-bearing surface, and is composed of a solid matrix and tissue fluid.

The most common knee injuries are found to be associated with damages to the articular cartilage.

Both intra-articular and extra-articular structures in the normal and abnormal knee joint could contribute to sounds generated during movement [30]. In the case of normal articular cartilage, it is not very likely that the joint surfaces generate measurable sound because the cartilage is smooth and slippery, and therefore creates a nearly frictionless surface during articulation. It is unlikely that appreciable sound would be generated from the flow of fluid between normal articular cartilage surfaces due to the smooth contact.

Abnormal structures and surfaces in the knee joint are more likely to generate sound during extension and flexion movements than normal structures. Softened articular cartilage, cracks, fissures, or thickened areas almost certainly increase the friction between the articular surfaces, and are therefore likely to increase the sounds emitted during normal joint movement [29].

When diagnosis of cartilage pathology is desired, a semi-invasive procedure such as arthroscopy (fiber-optic inspection and palpation of joint surfaces, usually requiring general anesthesia) is carried out [29]. But arthroscopy cannot be applied to patients whose knees are in a highly degraded state. Moreover, there is always the possibility that patients are found to have normal knees during arthroscopy. If a non-invasive technique could show the knee to be normal, diagnostic surgery could be avoided.

There exist a variety of non-invasive diagnostic methods that can detect damages to the knee. However most of the methods are unable to detect cartilage changes unless the changes are gross. Imaging techniques such as X-rays, computed tomography (CT), and magnetic resonance imaging (MRI) are all non-invasive, but they can capture only gross cartilage defects, and may not be useful for early detection of cartilage pathology [9]. In addition, they are not able to characterize the "functional integrity" (softening, stiffness, fissuring, etc.) of cartilage [29].

Due to the problems with arthroscopy and the limitations of imaging-based non-invasive techniques mentioned before, it is desirable to develop other non-invasive methods for early detection, localization, and classification of cartilage disorders.

It has been found that vibroarthrographic (VAG) signals, the recording of human knee joint vibration or acoustic signals during active movement of the leg, can be used as a non-invasive diagnostic tool to detect articular cartilage degeneration, and therefore determine the early joint degeneration or knee defects that are reflected in knee movements [31].

4.1.2 Vibroarthrographic (VAG) signals

Since abnormal knee joints are more likely to generate sounds than normal knee joints during an active movement of the leg. VAG signals, which record the vibration signals emitted from the knee joints, can be used to diagnose joint degeneration. Various methods to analyze and classify VAG signals have been developed to better interpret the signals, and realize efficient and accurate automatic diagnosis.

There are two important characteristics of VAG signals:
(i) Highly non-stationary in nature, because the quality of joint surfaces coming in contact may not be the same from one angular position (point of time) to another during articulation of the joint [9].
(ii) Multi-component due to the multiple sources of vibration.

Thus, VAG signals cannot be easily analyzed by common signal processing techniques such as the Fourier transform. Non-stationary signal analysis tools such as adaptive time-frequency methods are desired for the decomposition and analysis of VAG signals. Extensive work has been done to classify the abnormal VAG signals from normal ones, which can be used in fast screening and initial diagnosis to improve the efficiency and accuracy.

A reasonably large database of VAG signals of 89 human knee joints has been used to test and evaluate the classification algorithms. The signals were recorded by placing an accelerometer at the mid-patella position of the knee as the subjects swung the leg over an approximate angle range of $135° - 0° - 135°$ in 4s. The experimental protocol has been approved by the Conjoint Health Research Ethics Board of the University of Calgary. The VAG signal was pre-filtered and amplified before digitizing at a sampling rate of 2 kHz [9].

The database consists of 89 signals (51 normal and 39 abnormal). Previous works on the analysis and classification of this database using time-frequency methods are listed as follows.

4.1.3 Previous Works on VAG Classification

In [32], using only the autoregressive (AR) coefficients as discriminant features, the classification between normal and abnormal VAG signals provided an accuracy of 68.9% with the leave-one-out method. In [9], an adaptive time-frequency distribution (TFD) was constructed by minimum cross-entropy optimization of the TFD obtained by the matching pursuit decomposition algorithm. Parameters of VAG signals such as energy, energy spread, frequency, and frequency spread were extracted and each VAG was represented by a set of six features. An overall normal/abnormal screening accuracy of 68.9% was reached. Using Local Discriminant Bases (LDB) algorithm on the same database, the classification accuracies reached 76.4% [33].

4.2 Signal Decomposition and Feature Extraction

In this work, a parametric analysis method is used to screen abnormal knees from normal ones. Matching pursuit with Gabor dictionary is used to decompose the knee sound signals into time-frequency atoms. Atom parameters are analyzed to obtain discriminant information for 2-group classification. The same database of 89 VAG signals (51 normal and 38 abnormal), which was used in previous works, is employed again in the thesis work to test the proposed method and evaluate the classification performance of the extracted feature sets.

4.2.1 Signal Decomposition and Analysis

All the 51 normal and 38 abnormal VAG signals in the same database are plotted. One normal VAG and one abnormal VAG are randomly selected from the database and shown in Figure 4.1 and 4.2.

In order to look more into the characteristics demonstrated by the normal and abnormal

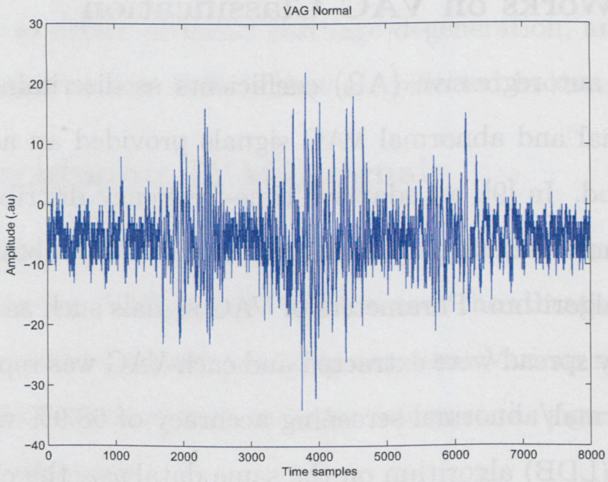

Figure 4.1: An example of normal knee sound signals. *au* - arbitrary unit. Sampling rate=2 kHz.

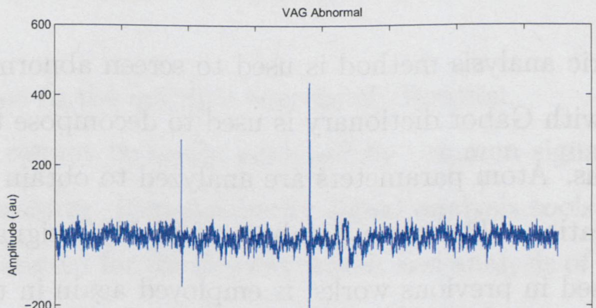

Figure 4.2: An example of abnormal knee sound signals. *au* - arbitrary unit. Sampling rate=2 kHz.

knee sound signals and define the discriminatory features for classification, the spectrum and spectrogram of the signals are also studied. Spectrum plots show the power spectral density. Spectrogram is the squared modulus of the short-time Fourier transform and is generally used to display the TF energy distribution over the TF plane. From the energy distribution on spectrograms, it is found the energy distribution is different for abnormal and normal knee sound signals. It is found that in general abnormal VAG signals have higher energy components, and show heavier distribution in the higher frequency range, It has been observed from the sample plots that in general normal knee sound signals tend to be more flattened, which is in line with the physical analysis of the normal and abnormal knee joints, that is, abnormal knee joints are more likely to generate sound and friction. Thus it is supposed that certain energy distribution over the frequency may be a useful feature for classification. Spectra and spectrograms of one normal VAG and one abnormal VAG randomly selected from the database are also plotted from Figure 4.3 to Figure 4.6 to show the visible difference between the normal and abnormal knee sound signals.

In [27], Umapathy et al. studied the matching pursuit decomposition algorithm and observed that the octave distribution can reflect the spectral similarities for the same category of signals . Since the abnormal and normal knee sound signals demonstrate different characteristics, it is expected that octave may carry distinguishing information to screen abnormal signals from the normal ones.

4.2.2 Parameter Analysis and Feature Extraction

Each knee sound signal in the database is decomposed with matching pursuit with Gabor dictionary. Since matching pursuit is a true non-stationary analysis tool, each knee sound signal (normal or abnormal) can be decomposed at one time and no signal segmentation are needed. In order to reduce the computational complexity of calculating all the inner products and search in the very redundant dictionary for the best atom, the number of pursuit iterations is pre-set to control the decomposition process, and local maxima is used to limit the searching area. The number of iterations of the pursuit is set to be 10,000.

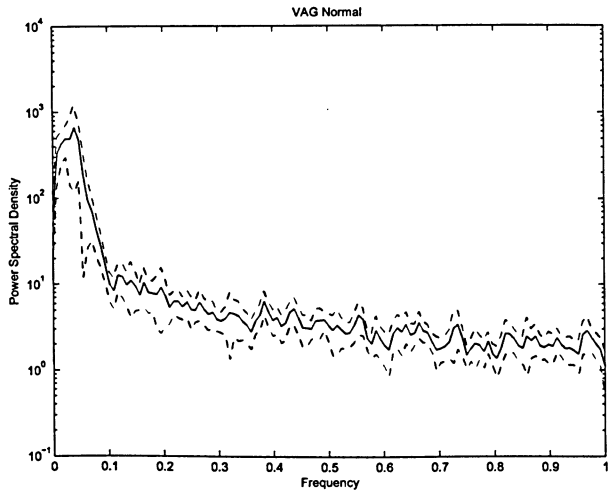

Figure 4.3: Spectrum of the normal knee sound signal in Figure 4.1. X-axis: normalized frequency, maximum frequency = sampling frequency/2.

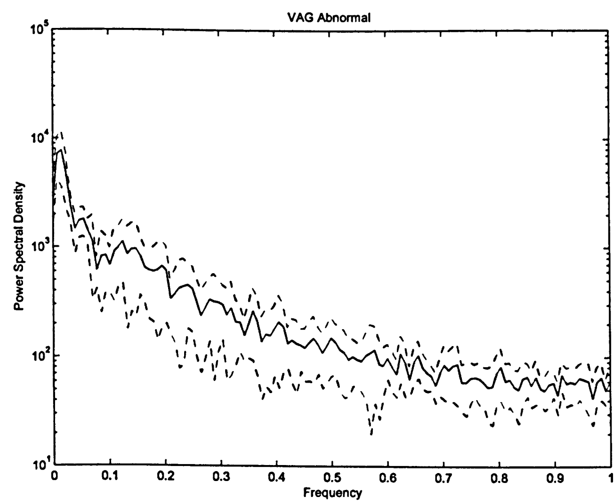

Figure 4.4: Spectrum of the abnormal knee sound signal in Figure 4.2. X-axis: normalized frequency, maximum frequency = sampling frequency/2.

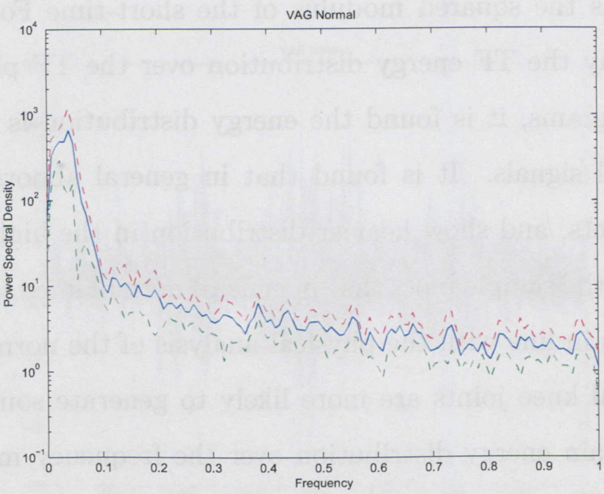

Figure 4.3: Spectrum of the normal knee sound signal in Figure 4.1. X-axis: normalized frequency, maximum frequency = sampling frequency/2.

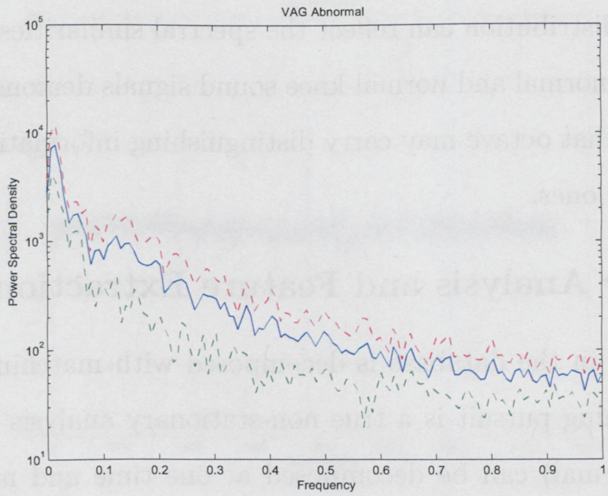

Figure 4.4: Spectrum of the abnormal knee sound signal in Figure 4.2. X-axis: normalized frequency, maximum frequency = sampling frequency/2.

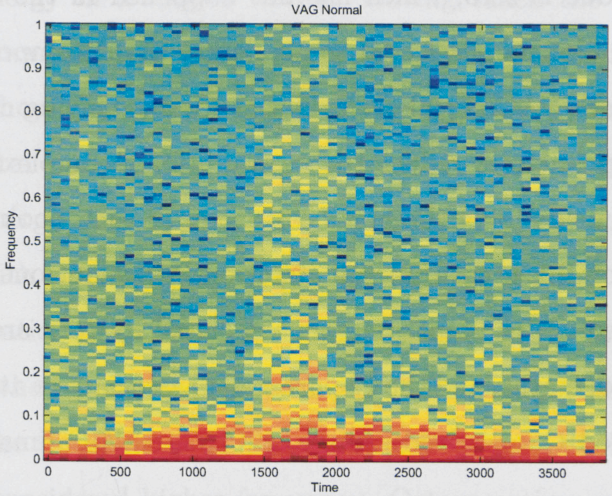

Figure 4.5: Spectrogram of the normal knee sound signal in Figure 4.1. X-axis: time samples. Y-axis: normalized frequency, maximum frequency = sampling frequency/2. Colors indicate different energy levels, with blue the lowest and red the highest.

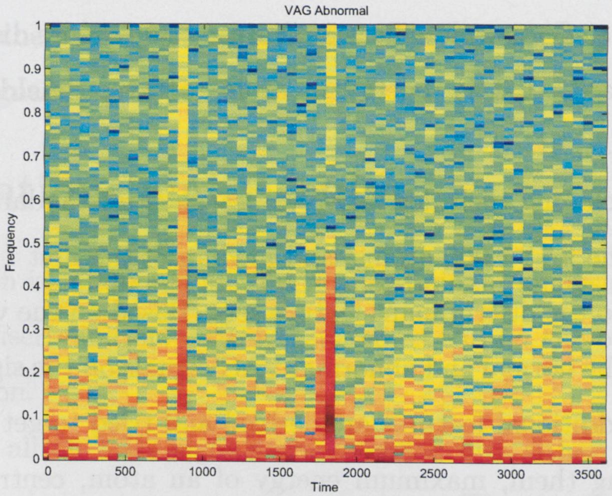

Figure 4.6: Spectrogram of the abnormal knee sound signal in Figure 4.2. X-axis: time samples. Y-axis: normalized frequency, maximum frequency = sampling frequency/2. Colors indicate different energy levels, with blue the lowest and red the highest.

the energy concentration status of the signal in the time-frequency distribution.

Central energy: energy in a "super" atom. Having taken the energy impact in each atom along the frequency axis into consideration, we assume there is one "super" atom at a frequency location whose energy can replace all the total energy in all the atoms and still reflects the actually total energy effect along the frequency axis. This "super" atom energy is defined as central energy in the paper. Thus, central energy is calculated as the sum of energy in each atom with the frequency weight divided by the total frequency.

Octave activeness ratio: the ratio of the sum of octaves in the lower frequency range to the sum of octaves in the higher frequency range. A frequency threshold is selected. Any frequency lower than the threshold is defined as lower frequency, while any frequency higher than the threshold is considered higher frequency. Octave activeness ratio is calculated as Octave activeness ratio shows in which range (lower or higher) the sum of octaves is larger. Since the octave distribution reflects the spectral similarities, this additional feature may reflect that in which range (lower or higher), the signal shows more activeness and higher energy distribution. From observation, abnormal VAG signals are more active in higher frequency range, thus the octave activeness ratio is supposed to be smaller than the ratio for the normal VAG signals.

4.3 Classifications and Results

Supervised classification is conducted and linear discriminant analysis (LDA) is employed to analyze and test the discriminatory feature sets, so as to find the optimum feature set for VAG signal classification.

In the experiment, all 89 VAG signals are classified into either normal class or abnormal class using the discriminant feature set under test. The classification performances of different feature sets are evaluated using the leave-one-out method to provide a least bias estimate. Since it is a 2-class supervised classification, only one canonical discriminant function is derived based on the selected features, and the histogram plots are used to show the classification performance. All the 8 parameters, including their derivative values such as standard deviation, mean, median of octave, innerProdR and innerProdI and realGG, have been studied and selected into the discriminant feature sets. Maximum atom energy, central energy, and octave activeness ratio are also computed and tried as discriminatory features. Different possible constructions of the discriminant feature sets are tried and tested. The optimum feature set is chosen based on the classification accuracy.

To illustrate the experiment process, the trials and results have been recorded in Table 4.1. The 2-group classification accuracy rates are all evaluated using the leave-one-out method, which is explained in Section 2.2.2. In Table 4.1, STD stands for standard deviation, Med stands for median, CE stands for central energy, ME stands for maximum atom energy, and OAR stands for octave activeness ratio.

After a long trying and comparing process, the optimum feature set, which brings up the best classification accuracy, is found to be: {octave activeness ratio, central energy, and standard deviation of innerProdI}. The classification performance of this feature set is illustrated as histogram plots in Figure 4.7 and 4.8. Among 51 normal knee sound signals, 37 are classified correctly while 14 are misclassified into abnormal class. Among 38 abnormal knee sound signals, 29 are classified correctly while 9 are misclassified into normal class. The overall accuracy reaches 74.2%.

4.4 Conclusions and Contributions

Since the same database is used in the above-mentioned previous works and the thesis work, a comparison of the classification performances is easy to conduct. By analyzing knee sound signals in time-frequency domain and reflecting their true non-stationary nature, a better result has been achieved than using the AR coefficients as discriminant features. While the same decomposition algorithm matching pursuit is used, the optimum feature set obtained in this work outperforms the feature set of energy, energy spread, frequency, and frequency spread. Although using LDB algorithm, the accuracy rate is higher than the rate obtained in the thesis work, the proposed techniques have lower computational complexity and a higher degree of automation, which increases the potential as a practical tool in machine

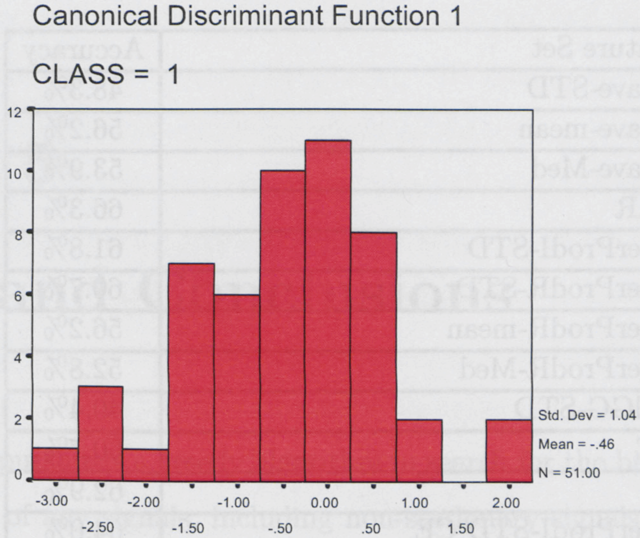

Figure 4.7: Knee sound signal classification: normal

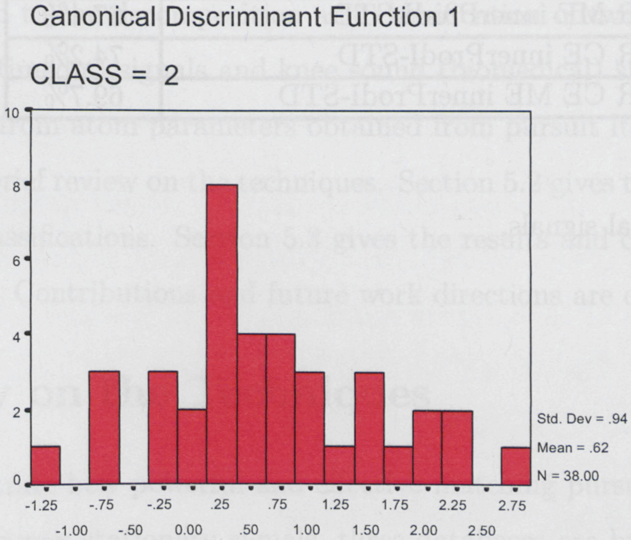

Figure 4.8: Knee sound signal classification: abnormal

Table 4.1: Classification performance of the feature sets in 2-group knee sound classification.

Feature Set	Accuracy
octave-STD	48.3%
octave-mean	56.2%
octave-Med	53.9%
OAR	66.3%
innerProdI-STD	61.8%
innerProdR-STD	60.7%
innerProdR-mean	56.2%
innerProdR-Med	52.8%
realGG-STD	40.4%
ME	60.7%
CE	62.9%
innerProdI-STD CE	64.0%
innerProdI-STD ME	58.4%
innerProdI-STD innerProdR-STD CE	62.9%
innerProdI-STD octave-mean CE	65.2%
OAR ME	68.5%
OAR CE	69.7%
OAR octave-mean	70.8%
OAR CE ME	73.0%
OAR ME innerProdI-STD	67.4%
OAR CE innerProdI-STD	74.2%
OAR CE ME innerProdI-STD	69.7%

classification of biomedical signals.

Chapter 5

Results and Conclusions

MATCHING pursuit is a greedy algorithm in search for the best time-frequency localized match of any signals, including non-stationary signals. At each iteration, the pursuit chooses at least one best matching atoms from a redundant dictionary. The properties of the signal components are explicitly given by the scale, frequency, time and phase indexes of the selected atoms. Matching pursuit algorithm can be used in adaptive signal representations, spectral estimation, compact signal coding, signal classification and compression, image representation, etc. In this work, matching pursuit algorithm with Gabor dictionary is applied to the decomposition and classification of two types of non-stationary signals: music (multimedia) signals and knee sound (biomedical) signals. Discriminant features are extracted from atom parameters obtained from pursuit iterations.

Section 5.1 is a brief review on the techniques. Section 5.2 gives the results and comments on music sample classifications. Section 5.3 gives the results and comments on knee sound signal classification. Contributions and future work directions are covered in Section 5.4.

5.1 Review on the Techniques

In order to demonstrate how powerful and effective matching pursuit is as a decomposition and analysis tool for non-stationary signals, three databases are built and classified in this work. The first two database contains music (multimedia) signals, and the last one is comprised of knee sound (biomedical signals). Both music and knee sound signals are highly non-stationary and multi-component in nature. The goal of the experiments is to analyze the non-stationary signals, find the discriminatory feature set, and classify all samples into different pre-set categories.

All the non-stationary signals in each database are decomposed into time-frequency atoms using matching pursuit with Gabor dictionary. Atom parameters are studied to find out the discriminant information. Good discriminating parameters are extracted and analyzed, and their derivative values, such as mean, median, and standard deviation, are also calculated and studied. Several additional features, such as central energy and octave activeness ratio, are also defined and derived. The atom parameters and their derivative values, along with the additional features, are selected and combined into various classification features sets. Since the group labels are pre-set for all the samples, supervised classification is conducted. All feature sets are fed to the linear discriminant analysis classifier (LDA). The classification accuracy rate is estimated using the leave-one-out method. The analysis and classification methodologies are the same for all three databases. However, since the physical characteristics are different for each group of signals, the numbers of pursuit iterations, the values of maxima, and the optimum discriminating feature sets are different for different databases, and the classification accuracy rates are different as well.

5.2 Conclusion on Music Classification

The first two databases are comprised of music samples. All the samples are selected and extracted from music CDs. The samples in use are single-channel recordings and the sampling rate is 44.1 kHz. The first music database contains 6 groups of samples, each sample having a duration of 5 seconds. The second database is comprised of 2 groups of samples, each sample having a duration of 10 seconds.

5.2.1 Results of Music Classification

The first experiment analyzes a database of 96 pieces of music samples. Each sample has the duration of 5 seconds. The 96 pieces of music samples equally belong to 6 groups, i.e.

christmas choir, country, greek music, jazz, rock, and scottish music. Each sample is decomposed with matching pursuit with Gabor dictionary. The number of the pursuit iterations is pre-set to be 1,000. For each iteration, a set of 100 maxima is selected to accelerate the decomposition. The optimum feature set, which brings up the best classification accuracy, is found to be: {standard deviation of octave, median of octave, standard deviation of innerProdI, standard deviation of realGG, and central energy}. The overall accuracy reaches 89.6%.

In the second experiment, the database is comprised of 112 pieces of music samples, 56 pieces belonging to rock-like class and 56 pieces belonging to classical-like class. Each music signal has the duration of 10 seconds, doubled in size from the first experiment. Each music signal is decomposed using matching pursuit with Gabor dictionary. Since the duration of each sample has doubled, the number of atoms under study has increased accordingly to reveal more discriminant information. Thus the first 2,000 atoms are extracted and analyzed. In order to accelerate the decomposition, for each iteration, a set of 300 maxima is selected. The optimum feature set is found to be: {standard deviation of octave in the first 2,000 atoms}. The classification accuracy is 100%, meaning all the samples are correctly categorized into classical-like or rock-like groups.

5.2.2 Comments on Music Classification

Music signals are highly non-stationary and multi-component in nature. In recent years, there have been many achievements on multimedia signal classification. Among them, some works aim on wave-formatted music classification. However, most of the existing techniques do not take into consideration the non-stationary behavior of the multimedia signals while deriving the discriminating features. Samples are examined in either the time or frequency domain where it is assumed that the signals are wide sense stationary. The computational complexity for most of the existing works is relatively high. And the classifications are mostly among farther-distanced sound groups, such as speech, music and noise, or advertisement, football and news. Only a few works analyze multimedia signals in joint time-frequency domain, using true non-stationary tools to extract discriminating features, where the classifications are among different music styles which is harder than distinguishing music from other sound recordings, such as, speech or noise.

In [10], Esmaili et al. proposed a technique using short-time Fourier Transform (STFT) where features are derived directly from the time-frequency domain. 143 music signals, with 5-second duration in each signal, are classified into six genres, that is, rock, classical, folk, jazz, pop, and country. Features extracted include entropy, centroid, centroid ratio, bandwidth, silence ratio, energy ratio, and location of minimum and maximum energy. LDA is applied to test the group classification of cases. The accuracy of classification reaches 92.3% using the leave-one-out method. The proposal deals with music signals in time-frequency domain, and features extracted reflect the non-stationary properties of music signals. The computational complexity is relatively low and classification accuracy is relatively high compared to previous works. However, since short-time Fourier Transform is used in this technique, music signals are still being segmented and the determination of optimum window size brings up challenges and uncertainty in practice.

In [27], Umapathy et al. also used matching pursuit, the same adaptive time-frequency decomposition algorithm employed in the thesis work, to analyze music samples. The music samples were treated as true non-stationary signals, and no segmentations were required. No window sizes need to be determined either. A database of 64 music samples each of 5-second duration was used. In the database, there are sixteen rock-like music, sixteen classical-like music, and 32 other types of music. All the music samples were decomposed into atoms, and atom parameter octave was used to create patterns based on a similarity measure of the music signals. These patterns were used to generate templates to classify the music signals into two different categories, i.e. rock-like music and classical-like music. All the sixteen of the actual rock-like music signals were correctly classified with 100% accurate, and fourteen out of the sixteen instrumental classical-like music signals were correctly classified with a classification accuracy of 87.5%. An overall correct classification accuracy reached 90%. Some important observations were also made, such as, the octave parameter obtained as a result of time-frequency decomposition exhibits potential discriminatory ability to classify audio signals, and the octave distribution reflects the spectral similarities for the same category of signals.

The thesis work uses the same time-frequency decomposition algorithm, matching pursuit, adopts some valuable observations in the work done by Umapathy et al., and further extends the analysis of atom parameters and the study of potential applications. In the second experiment of the thesis work, a classification between rock-like music and classical-like music is performed again, and the atom parameter octave is still used as the key discriminant feature. However, the database size in this work has increased, including 56 pieces of rock-like music samples and 56 pieces of classical-like music samples, and each sample has the duration of 10 seconds. Using standard deviation of octave as the discriminatory feature, the classification accuracy reaches 100%. In the first database of the thesis work, a 6-group classification is conducted. The classification performance of other atom parameters besides octave has been discovered and studied. It is found that standard deviation values of several other atom parameters besides octave also carry good discriminant information. It is observed that a combination of good discriminatory features may bring up improved results. It is also noted that adding more discriminatory features does not necessary improve the classification performance. When the accuracy rate reaches certain value, adding more features to improve the performance may become difficult. After the classification performance reaches the optimal, adding more features into the feature set may just reduce the accuracy rate.

The experiments on the music databases verify again that matching pursuit, as an adaptive time-frequency tool, decomposes non-stationary signals into atoms whose parameters contain good discriminant information for classification. The thesis work further proves that octave has the discriminatory ability to classify audio signals. It is also discovered that some other atom parameters besides octave carry satisfying discriminatory information as well. The derivative values of these parameters may act as good discriminant features, bringing good classification results. New feature central energy has good performance as well. The experiments and results may bring people's attention to more atom parameters besides octave, and inspire more interest in music classification using matching pursuit.

5.3 Conclusion on Knee Sound Classification

The third experiment deals with knee sound signals, also known as Vibroarthrographic (VAG) signal. VAG is the recording of human knee joint vibration or acoustic signals during active movement of the leg. It can be used as a non-invasive diagnostic tool to detect articular cartilage degeneration, and therefore determine the early joint degeneration or knee defects that are reflected in knee movements. Therefore, achieving automatic classification between normal knee sound and abnormal knee sound has great significance in biomedical world.

The VAG signals used in this work are pre-filtered (10 Hz to 1 kHz) and have a sampling rate of 2 kHz. The database under experiment is comprised of 89 VAG signals, in which 51 signals are emitted from normal knees and 38 signals are from abnormal knees.

5.3.1 Results of Knee Sound Classification

Again, matching pursuit with Gabor dictionary is used to decompose all the knee sound signals. The number of pursuit iterations is set to be 10,000. For each iteration, a set of 1000 maxima is selected to accelerate the decomposition. To reduce the computational complexity and achieve high efficiency, only the first 1000 atoms retrieved are analyzed. Several new features have been defined and derived from atom parameters. The optimum feature set, which brings up the best classification accuracy, is found to be: {octave activeness ratio, central energy, and standard deviation of innerProdI}. Among 51 normal knee sound signals, 37 are classified correctly while 14 are misclassified into abnormal class. Among 38 abnormal knee sound signals, 28 are classified correctly while 10 are misclassified into normal class. The overall accuracy reaches 74.2%.

5.3.2 Comments on Knee Sound Classification

Automatic classification of normal and abnormal VAG signals has its practical value in biomedical application. Several works have been done on the same database.

In [32], Krishnan et al. used only the autoregressive (AR) coefficients as discriminant features, and the classification between normal and abnormal VAG signals provided an accuracy of 68.9% with the leave-one-out method. In [9], an adaptive time-frequency distribution (TFD) was constructed by minimum cross-entropy optimization of the TFD obtained by matching pursuit. Parameters of VAG signals such as energy, energy spread, frequency, and frequency spread were extracted and each VAG was represented by a set of six features. An overall normal/abnormal screening accuracy of 68.9% was reached. Using Local Discriminant Bases (LDB) algorithm on the same database, the classification accuracies reached 76.4% [33].

Since the same database is used in the above-mentioned experiments and the thesis work, a comparison of classification accuracy rates is easy to conduct. By analyzing knee sound signals in time-frequency domain and reflecting their true non-stationary nature, a better result has been achieved than using the AR coefficients as discriminant features. While the same decomposition algorithm matching pursuit is used, the optimum feature set obtained in this work outperforms the feature set of energy, energy spread, frequency, and frequency spread. Although using LDB algorithm, the accuracy rate is higher than the rate obtained in the thesis work, the proposed techniques have lower computational complexity and a higher degree of automation, which increases the potential as a practical tool in machine classification of biomedical signals.

5.4 Contributions and Future Work

5.4.1 Summary of Thesis Work

The thesis work further proves that matching pursuit is a powerful decomposition tool for non-stationary signal analysis. To decompose music signals in higher frequency range (44.1

kHz) and knee sound signals in lower frequency range (2 kHz), the same algorithm is applied and the same Gabor atom dictionary is employed. The different natures and frequency ranges of samples don't require different decomposition methodologies, except for signals of larger size, more pursuit iterations are used and bigger maxima is selected to grasp more information from the signals and accelerate the decomposition process.

Matching pursuit is a true non-stationary method and no signal segmentations are required. Since for classification purpose, general characteristics of signals in a broad sense are sufficient, relatively few pursuit iterations and large maxima are always tried to accelerate the decomposition process. Therefore, the computational complexity is relatively low. It is observed several atom parameters carry good classification information. Of course, for different databases, atom parameters function differently. For example, standard deviation of realGG is a good feature in 6-group music classification, while it does not work well in 2-group knee sound classification. The optimum classification feature sets for different databases are different as well. However, in the three experiments, from matching pursuit atom parameters and the features derived from them, an optimum feature set can always be found which brings out good classification accuracy rate. In the three experiments, for non-stationary signals of different natures, high frequency or low frequency, small size or large size, multimedia or biomedical, 2-group classification or 6-group classification, matching pursuit with Gabor dictionary is always a powerful and efficient decomposition tool for signal analysis and classification.

5.4.2 Contributions

In time-frequency (TF) analysis, atoms are usually used for visualization in TF plane. The thesis is one of the very few works that analyze atoms statistically and extracts discriminant features directly from the parameters. Together with the similar works done by Umapathy et al. [27] and Esmaili et al. [10], the thesis work opens a door to the parametric analysis method in joint TFD.

As experiments for classification, the databases in use are relatively big, the sample sizes are relatively large, and the classification results are satisfying. Especially in the second experiment, music samples of 10-second in duration are firstly used, which doubles the size of the samples in peers' works [27, 10].

The parameters carrying strong discriminating information are unique, only achieved by matching pursuit in LastWave, The thesis work proves that matching pursuit is a powerful tool for non-stationary signal analysis and classification, and further study on atoms retrieved from matching pursuit and the associated parameters should be worthwhile and rewarding.

5.4.3 Future Directions

The atom parameters in the experiments are obtained from one type of implementation of matching pursuit with Gabor dictionary. The relationships between these parameters and the three basic atom parameters (scaling, translating and modulating) are not defined yet. More work could be done on parameter study, to understand the basic atom parameters and their various forms achieved in different implementations.

In addition, although the discriminant features extracted in the thesis work are strong for classifications, most of the features are abstract and do not carry clear physical meanings. Future work could be focused on better relating each parameter to the physical phenomenon of signals under research, so as to avoid random selections and combinations of classification features.

Bibliography

[1] L. Donagh, "A granular approach for the analysis of monophonic audio signals," *Proceedings of IEEE International Conference on Acoustics Speech Signal Process,* April 2003; 469-472.

[2] L. Cohen, "Time-frequency distributions - a review," *Proceedings of IEEE,* 1989; **77**(7):941-981.

[3] S. Krishnan and R. Rangayyan, "Automatic de-noising of knee-joint vibration signals using adaptive time-frequency representation," *Medical & Biological Engineering & Computing,* 2000; **38**(1):2-8.

[4] "Notes on Digital Signal Processing at Rice University," *http://www-dsp.rice.edu/software/optkernel.shtml.*

[5] L. Pazdera and J. Smutny, "Identification and analysis of acoustic emission signals by Cohen's class of time frequency distribution," it Proceedings of the 15th World Conference on Non-Destructive Testing in Rome, October 2000; 15-21.

[6] K. Umapathy, "Time-frequency modelling of wideband audio and speech signals," Master's thesis, Department of Electrical and Computer Enginnering, Ryerson University, Toronto, Ontario, Canada, 2002.

[7] G. Davis, "Adaptive nonlinear approximations," Ph.D. dissertation, Department of Mathematics, Courant Institute of Mathematical Sciences, New York University, New York, NY, U.S.A., 1994.

[8] S. Mallat, "*A Wavelet Tour of Signal Processing,*" Academic press, San Diego, CA, 1998.

[9] S. Krishnan, R. Rangayyan, G. Bell and C. Frank, "Adaptive time-frequency analysis of knee joint vibroarthrographic signals for noninvasive screening of articular cartilage pathology," *IEEE Transactions of Biomedical Engineering,* June 2000; **47**(6).

[10] S. Esmaili, S. Krishnan, K. Raahemifar, "Content based audio classification and retrieval using joint time-frequency analysis," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing,* May 2004; **5**(17-21):V-665-668.

[11] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Transactions on Signal Processing,* December 1993; **41**(12):3397-3415.

[12] G. Davis, S. Mallat and Z. Zhang, "Adaptive time-frequency approximations with matching pursuits," *Optical Engineering,* July 1994; **33**(7):2183-2191.

[13] E. Bacry, "LastWave Documentation,"
http://www.cmap.polytechnique/~bacry/LastWave/download_doc.html.

[14] K. Fukunaga, *Introduction to Statistical Pattern Recognition,* Academic Press, Inc., San Diego, CA, 1990.

[15] SPSS Inc., "SPSS advanced statistics user's guide," *User Manual,* SPSS Inc., Chicago, IL, 1990.

[16] "Lecture notes on the leave-one-out method,"
http://research.cs.tamu.edu/prism/lectures/iss/iss_l13.pdf.

[17] G. Tzanetakis, G. Essl and P. Cook, "Automatic musical genre classification of audio signals," *Proceedings of International Symposium on Music Information Retrieval (ISMIR),* 2001.

[18] C. Yang, "Music database retrieval based on spectral similarity," *Stanford University Database Group Technical Report,* 2001-14.

[19] E. Pollastri and G. Simoncelli, "Classification of melodies by composer with hidden Markov models," *Proceedings of IEEE First International Conference on Web Delivering of Music,* November 2001; **23-24**:88-95.

[20] H. Soltau, T. Schultz, M. Westphal and A. Waibel, "Recognition of music types," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing,* May 1998; **2**(12-15):1137-1140.

[21] E. Wold, T. Blum, D. Keislar and J. Wheaton, "Content-based classification, search, and retrieval of audio," *IEEE Multimedia,* 1996; **3**(3):27-36.

[22] M. Shan, F. Kuo and M. Chen, "Music style mining and classification by melody," *Proceedings of IEEE International Conference on Multimedia and Expo,* August 2002; **1**(26-29):97-100.

[23] Y. Zhang and J. Zhou, "A study on content-based music classification," *Proceedings of Seventh International Symposium on Signal Processing and Its Applications,* July 2003; **2**(1-4):113-116.

[24] S. Josephson, M. Schneider and D. Tamir, "A fuzzy expert system tool used for the analysis of musical timbres," *Proceedings of NAFIPS/IFIS/NASA 1st International Joint Conference of the North American Fuzzy Information Processing Society Biannual Conference,* December 1994; **18-21**:434 - 435.

[25] D. Pye, "Content-based methods for the management of digital music," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing,* June 2000; **6**(5-9):2437-2440.

[26] T. Lambrou, P. Kudumakis, R. Speller, M. Sandler, and A. Linney, "Classification of audio signals using statistical features on time and wavelet transform domains," *Pro-*

ceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, May 1998; **6**(12-15):3621 - 3624.

[27] K. Umapathy, S. Krishnan and S. Jimaa, "Audio signal classification using time-frequency parameters," *Proceedings of IEEE International Conference on Multimedia and Expo,* August 2002; **2** (26-29):II-249-252.

[28] G. Tzanetakis, "Manipulation, Analysis and Retrieval Systems of Audio Signals," Ph.D. dissertation, Department of Computer Science, Princeton University in Candidacy, NJ, U.S.A., 2002.

[29] S. Krishnan, "Adaptive filtering, modeling, and classification of knee joint vibroarthrographic signals," Master's thesis, Department of Electrical and Computer Engineering, The University of Calgary, Calgary, Alberta, Canada, 1996.

[30] W. Kernohan, D. Beverland, G. McCoy, A. Hamilton, P. Watson and R. Mollan, "Vibration arthrometry," *Acta Orthop. Scand.,* 1990; **61**(61):70-79.

[31] W. Kernohan, D. Beverland, G. McCoy, S. Shaw, R. Wallace, G. McCullagh and R. Mollan, "The diagnostic potential of vibration arthrography," *Clin. Orthop.,* 1986; **210**:106-112.

[32] S. Krishnan and R. Rangayyan, G. Bell and C. Frank, "Adaptive filtering, modeling and classification of knee joint vibroarthrographic signals for non-invasive diagnosis of articular cartilage pathology," *Medical & Biological Engineering & Computing,* November 1997.

[33] K. Umapathy and S. Krishnan, "Modified local discriminant bases and its application in signal classification," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing,* May 2004; **2**(17-21):II-745-748.

[34] D. Groutage and D. Bennink, "Feature sets for non-stationary signals derived from

moments of the singular value decomposition of Cohen-Posch (positive time-frequency) distributions," *IEEE Transactions on Signal Processing,* May 2000; **48**(5).

[35] G. Davis, S. Mallat and M. Avellaneda, "Adaptive greedy approximations," *Journal of Constructive Approximation,* 1997; **13**:57-98.

[36] G. Davis, S. Mallat and Z. Shang, "Adaptive time-frequency decompositions with matching pursuits," *Optical Engineering,* July 1994; **33**(7):2183-2191.

[37] L. Stankovic, "Time-frequency distributions with complex argument," *IEEE Transactions on Signal Processing,* March 2002; **50**(3):475-486.

[38] M. Goodwin and M. Vetterli, "Matching pursuit and atomic signal models based on recursive filter banks," *IEEE Transactions on Signal Processing,* July 1999; **47**(7):1890-1902.

[39] P. Durka, D. Ircha and K. Blinowska, "Stochastic time-frequency dictionaries for matching pursuit," *IEEE Transactions on Signal Processing,* March 2001; **49**(3):507-510.

[40] S. Jaggi, W. Karl, S. Mallat, and A. Willsky, "High resolution pursuit for feature extraction," *Applied Computational and Harmonic Analysis,* 1998; **5**:428-449.

[41] S. Krishnan, "Instantaneous mean frequency estimation using adaptive time-frequency distributions," *Proceedings of IEEE Canadian Conference on Electrical and Computer Engineering,* 2001; **1**(1):141-146.

[42] G. Guo, H. Zhang and S. Li, "Boosting for content-based audio classification and retrieval:an evaluation," *Proceedings of IEEE International Conference on Multimedia and Expo,* August 2001; **22-25**:997-1000.

[43] G. Lu and T. Hankinson, "A technique towards automatic audio classification and retrieval," *Fourth International Conference on Signal Processing,* October 1998; 1142-1145.

[44] G. Tzanetakis and P. Cook, "MARSYAS:a framework for audio analysis," *Organized Sound,* 2000; 4 (3).

[45] G. Tzanetakis and P. Cook, "Multifeature audio segmentation for browsing and annotation," *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,* 1999; 103-106 .

[46] J. Foote, "Content-based retrieval of music and audio," *Multimedia Storage and Archiving System II, Proceedings of SPIE,* 1997; 138-147.

[47] J. Foote, "An overview of audio information retrieval," Multimedia Systems, ACM Press/Springer-Verlag, January 1999; **7** (1):2-11.

[48] K. Umapathy, S. Krishnan, V. Parsa and D. Jamieson, "Discrimination of pathological voices using an adaptive time-frequency approach," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing,* May 2002; **4**(13-17):IV-3852-3855.

[49] K. Umapathy, S. Krishnan and G. Sinanian, "A General perceptual tool for evaluation of audio codecs," *Proceedings of IEEE Canadian Conference on Electrical and Computer Engineering,* May 2003; **2** (4-7):II-683-686.

[50] K. Umapathy, S. Krishnan, V. Parsa and D. Jamieson, "Time-frequency modeling and classification of pathological voices," *Proceedings of EMBS/BMES 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society,* 2002; **1**:116-117.

L. Donagh, F. Bimbot, and W. Gribonval, "A granular approach for the analysis of monophonic audio signals," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing,* April 2003; **6**(6-10):VI-469-472.

[51] L. Lu and H. Zhang, "Content analysis for audio classification and segmentation," *IEEE Transactions on Speech and Audio Precessing,* October 2002; **10**(7):504-516.

[52] L. Lu and H. Zhang, "A robust audio classification and segmentation method," *ACM Multimedia,* 2001; 203-211.

[53] S. Pfeiffer, S. Fischer and W. Effelsberg, "Automatic audio content analysis," *ACM Multimedia,* November 1996; 21-30.

[54] T. Li, M. Ogihara and Q. Li, "A comparative study on content-based music genre classification," *Proceedings of ACM SIGIR 26th Annual International Conference on Research and Development in Information Retrieval,* 2003.

[55] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing,* July 2002; **10**(5):293-302.

[56] W. Chai and B. Vercoe, "Folk music classification using hidden Markov models," *Proceedings of International Conference on Artificial Intelligence,* 2001.

[57] Z. Liu, J. Huang, Y. Wang and T. Chen, "Audio feature extraction & analysis for scene classification," *IEEE Workshop on Multimedia Signal Processing,* June 1997; 343-348.

[58] H. Steil, J. Kirchgessner, V. Moscardo and A. Mitteregger, "Survey of medical data from data acquisition to high quality patient treatment," *Proceedings of IEEE Conference on Information Technology Applications in Biomedicine,* April 2003; 378-381.

[59] J. Lee, C. Jiang and T. Yuan, "Vibration arthrometry in patients with knee joint disorders," *IEEE Transactions on Biomedical Engineering,* August 2000; **47**(8):1131-1133.

[60] M. Chu, I. Gradisar and L. Zavodney, "Possible clinic application of a noninvasive monitoring technique of cartilage damage in pathological knee joints," *Journal of Clinical Engineering,* January-March 1978.

[61] W. Kernohan and R. Mollan, "Microcomputer analysis of joint vibration," *Journal of Microcomputer Applications,* 1982; **5**:287-296.

[62] S. Esmaili, "Content based audio watermarking and retrieval using time-frequency analysis," Master's thesis, Department of Electrical and Computer Engineering, Ryerson University, Toronto, Ontario, Canada, 2004.

[63] S. Quadri, "Indexing of American football video using MPEG-7 descriptors and MFCC features," Master's thesis, Department of Electrical and Computer Engineering, Ryerson University, Toronto, Ontario, Canada, 2004.