Universidade Nova de Lisboa
Faculdade de Ciências e Tecnologia
Departamento de Informática

Automatic Cymbal Classification

Hugo Almeida, nº 26522

Dissertação apresentada na Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa para a obtenção do grau de Mestre em Engenharia Informática

Orientadora
Prof(a). Doutora Sofia Cavaco

Lisboa

Novembro de 2010

Nº do aluno: 26522
Nome: Hugo Ricardo da Costa Almeida

Título da dissertação:
Automatic Cymbal Classification

Keywords:

- Automatic Classification
- Cymbal Classification
- Music Classification
- Music Information Retrieval (MIR)
- Drum Kit
- Cymbals
- Information Theoretic Algorithms
- Principal Component Analysis (PCA)
- Independent Component Analysis (ICA)
- Non-Negative Matrix Factorisation (NMF)
- Sparse Coding
- Non-Negative Sparse Coding
- Independent Subspace Analysis (ISA)
- Sub-band Independent Subspace Analysis (Sub-band ISA)
- Locally Linear Embedding (LLE)
- Prior Subspace Analysis (PSA)

Resumo

A maioria da investigação que acenta sobre transcrição automática de música, foca-se primariamente nos instrumentos de tom definido como a guitarra e o piano. Ao contrário destes últimos, instrumentos de tom indefinido, tal como a bateria, que é uma colecção de instrumentos deste tipo, têm sido muito desconsiderados. No entanto, ao longo dos últimos anos e provavelmente devido à sua popularidade no panorama musical ocidental, este tipo de instrumento começou a gerar um maior nível de interesse.

O trabalho relacionado com a transcrição automática da bateria foca-se principalmente na tarola, bombo e prato de choque. No entanto, muito é o trabalho que necessita de ser realizado com o intuito de efectuar transcrição automática de todos os instrumentos de tom indefinido. Os pratos da bateria são um exemplo de um tipo de instrumentos de tom indefinido e com características acústicas particulares, sobre o qual não tem recaído muito atenção por parte da comunidade cientifica.

Uma bateria contém vários pratos que usualmente ou são tratados como se fossem um instrumento único ou são ignorados pelos classificadores de instrumentos com tom indefinido. Propomos preencher esta lacuna e como tal, o objectivo desta dissertação é a classificação automática de pratos de bateria e a identificação das classes de pratos a que pertencem. Conseguimos preencher esta lacuna dando uso a dois algoritmos - um da área de teoria de informação e outro de classificação, os quais serão descriminados e explicados em capítulos vindouros.

Os pratos de bateria apresentam muitas similiridades, que vão desde a sua geometria, material de que são feitos, características sonoras, até às características espectrais. Os testes que são executados sobre instrumentos da bateria, na sua maioria, usam instrumentos muito diferentes

entre si, como o bombo, a tarola e o prato choque. Assim, a grande vitória deste trabalho encontra-se na obtenção de classificações correctas de diferentes pratos de bateria, tendo em atenção que existe um maior grau de dificuldade neste caso, dadas as similiridades entre os intrumentos testados.

Abstract

Most of the research on automatic music transcription is focused on the transcription of pitched instruments, like the guitar and the piano. Little attention has been given to unpitched instruments, such as the drum kit, which is a collection of unpitched instruments. Yet, over the last few years this type of instrument started to garner more attention, perhaps due to increasing popularity of the drum kit in the western music.

There has been work on automatic music transcription of the drum kit, especially the snare drum, bass drum, and hi-hat. Still, much work has to be done in order to achieve automatic music transcription of all unpitched instruments. An example of a type of unpitched instrument that has very particular acoustic characteristics and that has deserved almost no attention by the research community is the drum kit cymbals.

A drum kit contains several cymbals and usually these are treated as a single instrument or are totally disregarded by automatic music classificators of unpitched instruments. We propose to fill this gap and as such, the goal of this dissertation is automatic music classification of drum kit cymbal events, and the identification of which class of cymbals they belong to.

As stated, the majority of work developed on this area is mostly done with very different percussive instruments, like the snare drum, bass drum, and hi-hat. On the other hand, cymbals are very similar between them. Their geometry, type of alloys, spectral and sound traits shows us just that. Thus, the great achievement of this work is not only being able to correctly classify the different cymbals, but to be able to identify such similar instruments, which makes this task even harder.

Index

Figures Index

Tables Index

1. Introduction

Music is constantly present in our everyday activities. From the first second of our day when we wake up with the radio from our alarm clock, to the most common entertainment mediums like cinema, television, video games, and of course radio, to the music we ear and sing while bathing or while traveling to work. It is quite amazing how in just a few seconds after the start of a song we are able to recognize and identify it. However, recognition of a song or a piece of music does not enable a listener to transcribe it.

Transcription is the ability to identify and register instruments', harmonic[1], rhythmic[2], and melodic[3] features of a piece[4] of music, using standard staff notation[5]. It requires the attainment of aural skills[6] and music theory knowledge and comprehension, which are only possible through training and study. To achieve a level of proficiency in transcription that is fast and accurate can take a long time. This way, for a beginner, several weeks may be required to transcribe only one of the instruments from a musical piece, without guaranties of total accuracy. Although not deprived of usefulness this ability enables little utilizations besides transcription and music composition.

1 Harmony deals with pitches that are played at the same time [Burrows 99]. The pitch of a note can be defined scientifically in terms of its sound waves frequencies. Similarly in music, a pitch is a fixed sound which can be identified using a series of letters ranging from A to G. So, every note you hear from a musical instrument has its own pitch [Burrows 99]. When at least three different notes sound together in the same instrument, the resulting effect is a chord [Burrows 99].

2 A pulsing effect that we feel when listening to a piece of music [Burrows 99]; usually its main engines are the percussion instruments.

3 Melody refers to the deliberate arrangement of series of pitches – what most people would call a tune [Burrows 99].

4 Throughout this thesis, music, song, and piece will be used interchangeably, refereeing to the same thing.

5 Staff notation consists of the written representation of all rhythmic, harmonic, and melodic elements in a piece of music. The notation is written in five lines which are known as the staff [Gerou 96].

6 Hearing and sigh-reading skills.

If extended to a computer system (automatic) music transcription can be a very useful asset. It can be used in computerized music education as a learning aid for people wishing to learn how to play a piece of music where there is only access to an audio recording, and not the necessary skills to attempt transcription themselves. Areas of entertainment such as karaoke [Ryynänenm 08], music composition [Simon 08], and even song data base retrieval through humming - known as query by humming [Ghias 95], are some of the other potential applications.

Automatic music transcription (AMT) is a very hard problem to tackle, mainly due to representation issues. These are a result of music's many complex structures, which are a combination of mathematical (harmony, rhythm, and melody), and non-mathematical (tension, expectancy, and emotion) variables. Hence, computerized representations of these variables, along with the transformations used in audio processing, add even more to the complexity of this area [Dannenberg 93]. The number of note sources targeted for transcription, and the number of notes played at the same time are also detrimental to the accuracy of a transcription. When notes are played one at a time we are in the presence of monophonic music. On the other hand, if there is more than one note being played like a chord or when more than one instrument plays a note at the same time, we are in the presence of polyphonic music. Both monophonic and polyphonic transcription can be handled in a single or multiple instrument environments.

Salience, perception, pitch matching, complexity of a piece of music, and overlooking rhythm are discussed in [Byrd 02] as some of the most common problems of monophonic and polyphonic music regarding music information retrieval (MIR) for pitched instruments. A great deal of research on AMT is usually focused on pitched instruments. FitzGerald gives some possible justifications regarding the preference for this type of instruments [FitzGerald 04]:

> *This is perhaps as a result of the predominantly melodic and harmonic based nature of most of Western Art music and of Western popular song as opposed to the more rhythmic based musical traditions such as that of Indian tabla playing and much of the music of*

Africa. It is also perhaps as a result of a feeling that the harmonic series of partials that go to make up a given pitch are easier to model than the noisy frequency spectra associated with most drum sounds.

However, over the last few years indefinite pitched instruments, mainly percussion instruments, started to garner more attention. From these, the one that stands out the most is the drum kit (see chapter 3.1 for more details on this instrument[7]), especially because of its increasing popularity in western music landscape. This growth in interested by the scientific community is also due to its usefulness in a great variety of musical situations where AMT is needed. Query by beat boxing [Kapur 04] is one of them, it's an information retrieval method for music databases based on the same concept as query by humming, but seen primarily as applicable for Disk Jockey (DJ) usage. AMT of drum kit events can also be used as an aid for people wishing to transcribe the drum kit parts played in a song, or for studying this instrument. Producers and music lovers can also gain from the development of tools based upon AMT of the drum kit. If an audio recording has enough quality the drum track can be sampled[8] to be used in other musical pieces. It is also possible to organize libraries of drum samples and drum loops by type of beat, tempo, or genre. Users with an enormous database of music could organize them by musical style based on the type of drum parts detected. Since some of the existing genres have a much defined rhythm structure, it is possible to label them based on that. Therefore, there is a whole world of new possibilities for the musician, the producer, and even for the everyday music enjoyer with AMT of drum kit events.

Most of the work on automatic drum transcription is focused on combinations of snare drum [Tindale 04], hi-hat, and bass drum (also known as the kick drum, these two names will be used interchangeably throughout the text) [Paulus 06, FitzGerald 06], which are the main

7 A drum kit is a collection of percussion instruments, so it is not accurate to call it an instrument. For simplicity and also because it is of common usage, and seeing that this issue is not relevant, in this dissertation besides drum kit we will also refer to it as an instrument and drum set.

8 In music, sampling is the process of recording a sound source one part at a time. Typical parts (samples) include each note recorded from a musical instrument [Sam 08], or in the case of a drum kit, each hit on its various instruments. A small part of a song can also be sampled in its entirety, or just one of the instruments. The use of this technique is a very common practice in Hip-Hop.

instruments of a drum kit. To the best of our knowledge, transcription of elements like the open hi-hat or even the different cymbals has been neglected. Yet, an accurate transcription of drum kit events will never be possible without the transcription of different types of cymbals, and in the case of a hi-hat, if it is open, closed, or half-open (just to name a few possible uses of this instrument). The goal of this dissertation is to fill this void. Here we explore automatic cymbal classification[9] and the identification of which class of cymbals the cymbal played belongs to. Classification is part of the transcription process. To perform correct transcription we have to first identify what instruments are being played, following this with detection of its positioning in the piece of music. We will focus on the five most used types of cymbal classes – crash, ride, splash, china, and hi-hat (for more information on each of these classes check chapter 3.2). Our study will only regard monophonic events from two or three cymbals played consecutively. Even though this work will only regard cymbal events, a great deal of issues will arise. From capturing all the dynamic nuances played by the drummer (strong or weak hits), classification of up to three cymbals played consecutively, to cymbals with different sizes, shapes, and timbres, these are some of the characteristics that will drastically increase the complexity of the work developed. Still, another problem arises from the typical harmonic series found in this type of instrument – it is harder to accurately classify a cymbal do to its noisy frequency spectra.

To steer our work in a good direction we chose to apply the cornerstones of the majority of information theory algorithms (IFA) – Principal Component Analysis (PCA) [Cavaco 07] [FitzGerald 04], Independent Component Analysis (ICA) [Abdallah 03] [Cavaco 07] [FitzGerald 04], and Non-Negative Matrix Factorization (NMF) [Smaragdis 03] [Hélen 05] [Moreau 07] [Virtanen 07], for sound source separation, combined with a classification algorithm for disclosing to what cymbal each sound sample pertains to. As we had predicted, PCA due to its constraints did not give satisfactory results. ICA's results were also not very satisfactory, so we decided to focus our attention on NMF. This algorithm was chosen because of encouraging results when used as a standalone technique, as seen on [Smaragdis 03] and [Virtanen 07]. With NMF we were able to achieve a great level of success by

9 From this point on, automatic classification will be simply designated as classification.

accurately classifying various combinations of two cymbals played sequentially, while with three cymbals the results were also very good, as with two cymbals.

We will start our journey by overviewing a collection of introductory topics. These range from the physical behavior of sound (chapter 2); physical characteristics and behavior of cymbals, and drum kit description (chapter 3). Afterwards, analysis and exploration of previous work will ensue with chapter 4 - State of the Art. There, we review several algorithms, their pros, and cons and possible applications to the problem at hand. Next, in the fifth chapter, we explain in detail the proposed system to solve our problem. This document will conclude with the analysis of the results on chapter 6 – Results and Discussion, and with the conclusions and future work on chapter 7.

This work was used as the basis for a paper with the title – Automatic Cymbal Classification Using Non-Negative Matrix Factorization, written by Hugo Almeida and Sofia Cavaco, and submitted to an international conference.

2. The Physics and Math of Sound

As the reader may be aware of, for the development of a work of this magnitude a high level of study and research is needed. Thus, we start by reading the ones that preceded us, those who strived to success that paved the way. Through papers and thesis we are introduced to a new and very scientific world, with a whole new jargon for us to cope with, with a whole new set of rules. With all this in mind we will try our best to achieve the type of approach portrayed in [Eco 98]:

> *Once decided for whom to write for (for all mankind and not just for the evaluator) it is essential to decide how to write*[10].

We will write this thesis with one objective in mind, to always try to clearly explain all its content, independently of the level of knowledge of the reader. Thus, in an effort to elaborate a very comprehensive source of knowledge we will start by taking a look at how sound behaves and how digital systems can capture and mathematically represent sounds. If the reader is knowledgeable about the subjects studied in this chapter, he/she is free to jump over to the third chapter of this dissertation.

2.1. From Sound Wave to Waveform

Have you ever wondered how it is possible for a sound to travel from a speaker to your ears? Figure 2.1 is an illustration of what ensues, since a sound is emitted by a pair of speakers until it reaches our ears. The dots in the picture represent air molecules. The regions with great density of molecules are called areas of compression - where the air pressure is greater than the one from the atmosphere. On the other hand, the dispersed dots are areas of

10 This is a translation from the portuguese version of [Eco 98]: Uma vez decidido para quem se escreve (para a humanidade e não para o relator), é necessário decidir como se escreve.

rarefaction, regions where pressure is lower than the one exerted by the atmosphere. The small arrows in the diagram represent the movement of a sound wave through a channel, which is created by a translation of the compressed area inwards, as opposed to the outwards movement of the scattered air molecules [Everest 01].

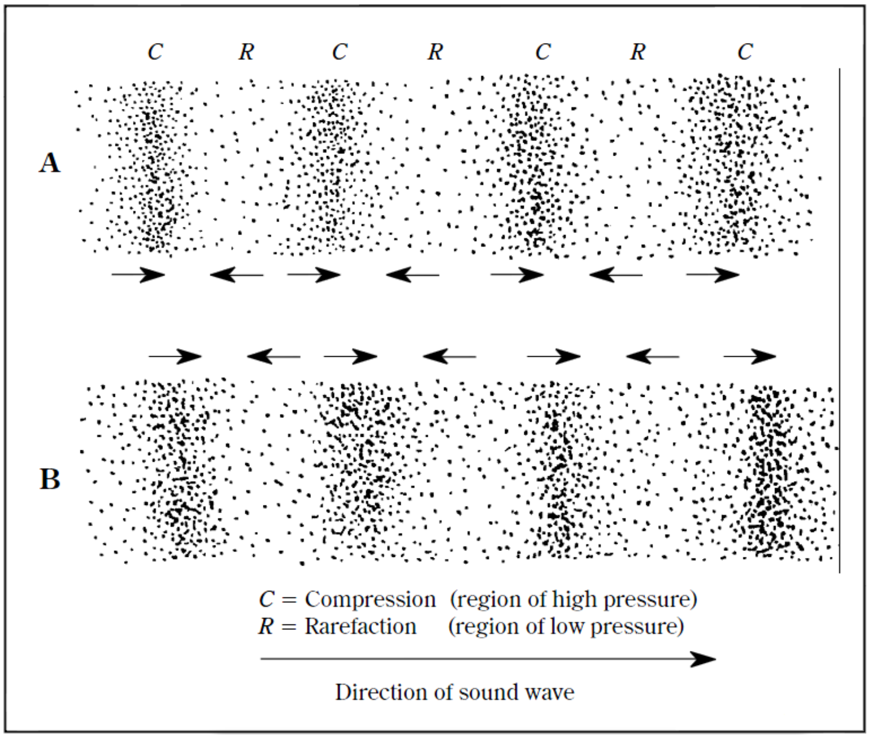


Figure 2.1 – From [Everest 01], the effect of sound pressure on air molecules.
(A) – Sound pressure is responsible for air particles being pressed together in some regions, and sparse in others
(B) – A small movement of the sound wave from the position occupied in A to a new one.

For the sound to be able to transit along the air, two conditions have to be met; first, there has to exist an equilibrium position to which the air molecules may be able to return to after compression or rarefaction; and secondly, the force that tends to push the air molecules back to equilibrium has to be proportional to the distance traveled [Berg 95]. So, air pressure tends to equilibrium, i.e., atmospheric pressure. A speaker develops an augmentation in the air pressure when it discharges the first sound wave. This establishes regions of compression (areas of the picture were the arrows are pointing to the right), and by extension, areas in the air with low pressure (areas of the picture were the arrows are pointing to the left). The collisions between particles near the speaker have two effects - restore the particles near the speaker to equilibrium, and displace the neighboring particles, which will enable the sound waves to move along the air. These movements are responsible for making the sound waves travel through the ear channel, which introduces changes in the wavelength of the sound wave. The end result of this is our perception of sound.

Now let us suppose that instead of reaching our ears the sound waves reach a microphone connected to a computer. In this particular case the information traveling in the sound waves will have to be digitized so it can be interpreted by a computer. When it comes to convert them to a digital medium their continuous information (in nature these waves are analog) will have to be transposed into discrete values. The digital and mathematical representation of the sound wave is called waveform, and is illustrated in figure 2.2 – B. This consists of representing the displacement of the air particles through time. In figure 2.2 we see the relationship between air pressure and the mathematical representation of a sound wave, where for example, values of compression represent high amplitude amounts. Now let us take a look at how the sound waves are translated into waveforms.

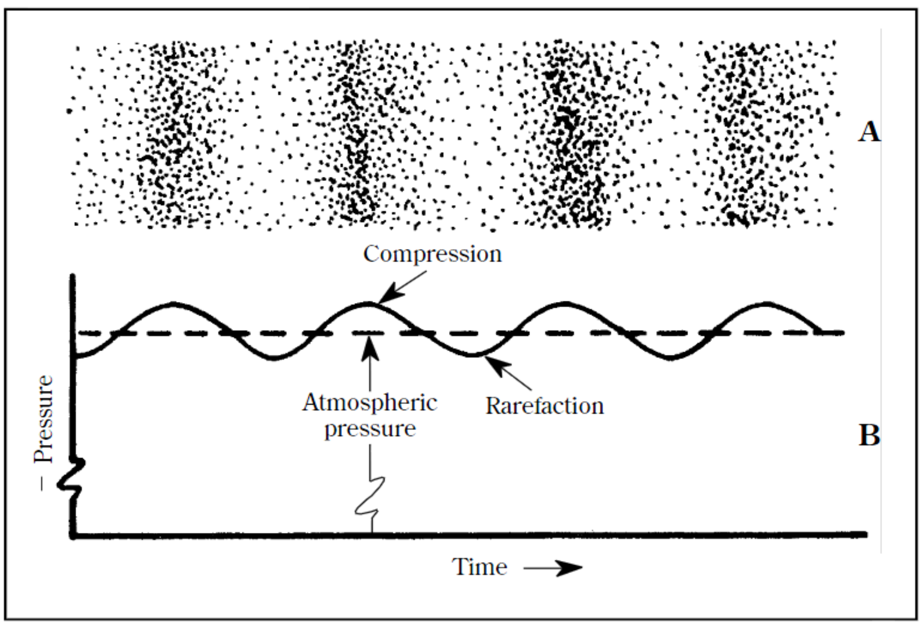


Figure 2.2 – From [Everest 01], relationship between a wave form (B) and the pressure values in the air (A).

Audio digitization systems use time sampling and amplitude quantization to encode the infinitely variable analog signal as amplitude values in time [Pohlmann 00]. Samples are taken at irregular intervals from an analog signal to create a discrete signal. The number of samples recorded per second is known as sampling frequency.

This is enough to guaranty the reconstruction of a signal with the same frequency as the original one, if the sampling theorem is taken into consideration. This theorem defines the relationship between the analog signal and the sampling frequency, specifying that the sampling frequency must be at least twice the highest signal frequency in order to allow reconstruction of the signal. More specifically, audio signals containing frequencies between 0 and $S/2$ Hz (Nyquist frequency) can be accurately represented by a sampling frequency of S samples per second [Pohlmann 00].

Figure 2.3 is a good visual example of what happens in the time sampling stage if the sampling theorem is followed. The samples will contain the same information as the original signal. Thus, the signal is reconstructed without loss of information [Pohlmann 00]. If the sampling theorem is not respected, information from the original signal will be lost, and it will not be possible to have the original signal reconstructed accurately in the discrete signal [Pohlmann 00]. As you can see in figure 2.4, the sampling frequency (44 kHz) is not two times the frequency of the analog signal (36 kHz) (figure 2.4 - A). This will in turn originate a deficient sampling frequency (figure 2.4 – B) blocking any possibility of an accurate reconstruction of the analog signal into a discrete one (figure 2.4 – C).

Since the machine representation of amplitude is limited by the number of bits used, the amplitude of each sample must be quantized, that is, the actual amplitude of the sample is rounded to be converted to a k bit number. Because amplitudes can have a high number of decimal values, if k is small, more quantization errors can be produced [Widrow 61].

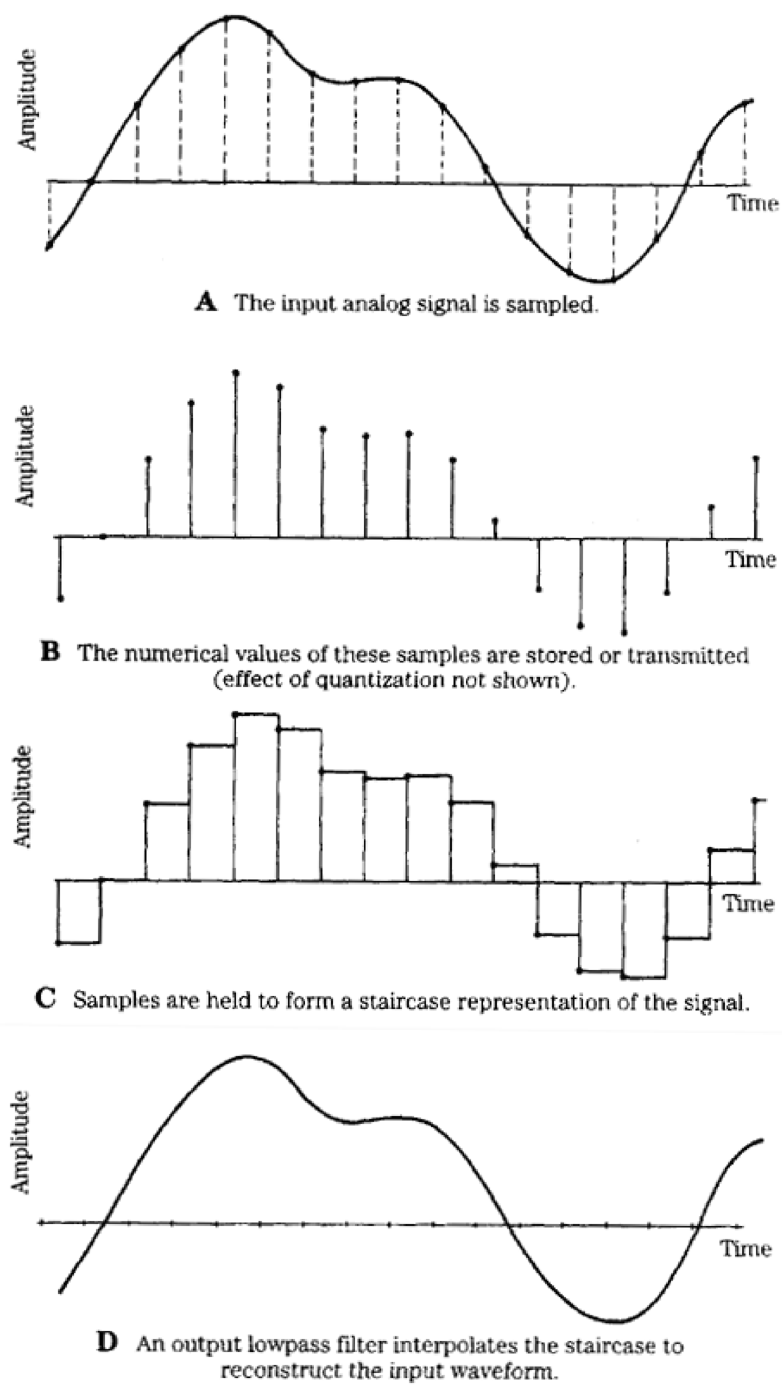


Figure 2.3 – From [Pohlmann 00], the effect of time sampling on an analog signal.

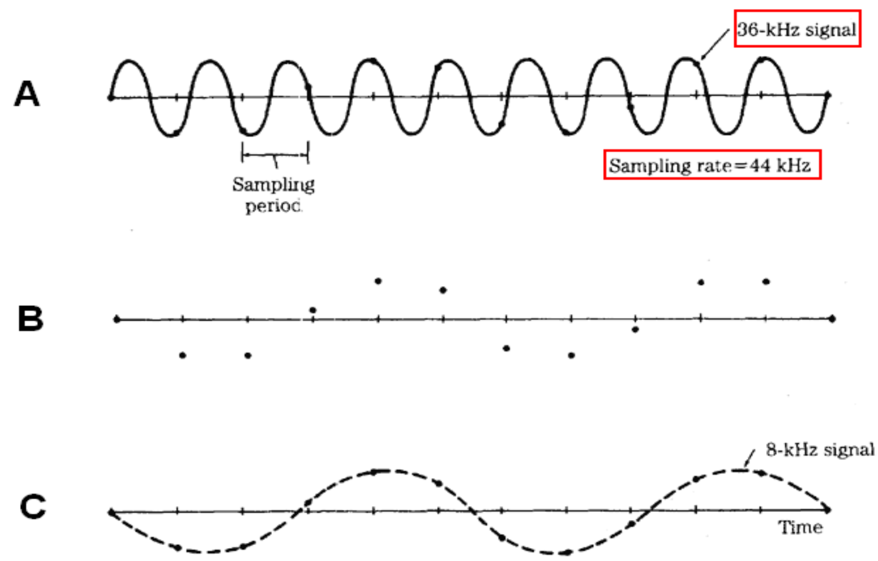


Figure 2.4 – From [Pohlmann 00], the effect of not obeying the sampling theorem.
(A) – The orginal signal. (B) – The stored samples.
(C) – The inaccurate representation of the reconstructed signal.

2.2. Spectrograms

After the sound has been digitized into a computer it is possible to perform operations that enable a better retrieval of information for analysis. One of these operations is known as Fourier Transform (FT), a mathematical tool that enables decomposing time signals (such as waveforms) into the frequency domain. The discrete Fourier transform (DFT) is used instead of the FT to obtain a sampled spectrum for discrete time signals of finite duration. Just as the FT generates the spectrum of a continuous signal, the DFT generates the spectrum of a discrete signal expressed as a set of related sinusoids. The DFT takes samples of a waveform and operates on them as if they were an infinitely long waveform comprised of sinusoids [Pohlmann 00]. So with DFT it is possible to demonstrate that a sound input may be described as the combination of various other sinusoids. Nonetheless, the DFT is not a very efficient computational technique when compared to fast Fourier transform (FFT) [Burrus 08], so this last one is used instead.

Applying FFT to the input signal might not be enough to gather detailed information on the signal's attributes, simply because we lose information on its temporal variations. Very rarely do we hear natural sounds with a constant value of frequency through time, and as such we must use another method to better analyze their time-varying frequency content. To do so the input signal can be divided in windows with a time based function performing FFT on each one of the windows. This technique is named Short-Time Fourier Transform (STFT) which specifies magnitude versus time and frequency for any signal [Cohen 95]. Even though the FFT (and consequently the STFT) also give information about the initial phase of the frequency components of the waveform, here we will not make use of this information; we will only use the magnitude information.

A windowing function is illustrated in figure 2.5. This signal is broken into chunks that are multiplied by the windowing function, which is embodied by the series of red curves that are applied to the signal being analyzed and represented in blue. Afterwards, the results of applying FFT to each window can then be placed together in a single matrix called a spectrogram, which is a graphical display of the magnitude of STFT.

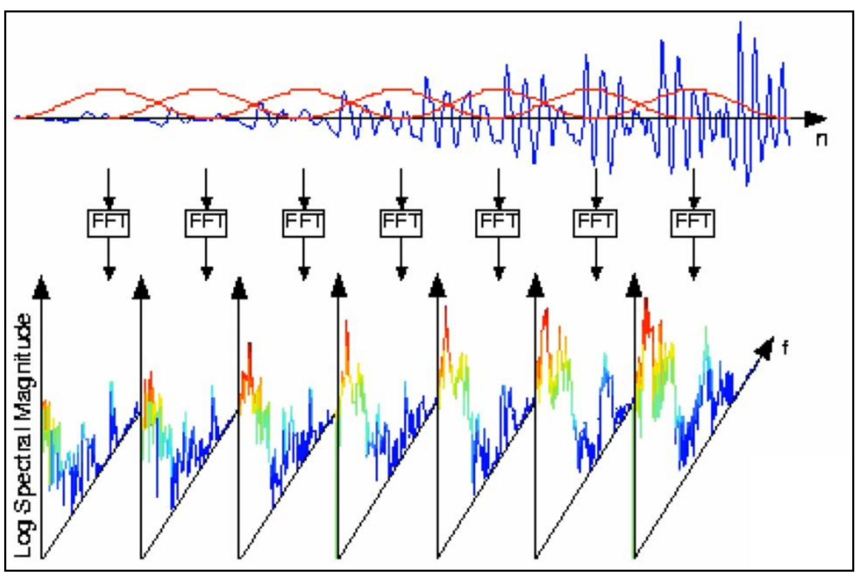


Figure 2.5 – From [ECE 10], short-time Fourier transform.

In equation 2.1 the spectrogram is represented in matrix S where a_{ft} is an amplitude value at time frame $t \epsilon \{0, \ldots, T\}$ and frequency bin $f \epsilon \{0, \ldots, F\}$. In this following example (figure 2.6) the magnitude of the frequency components of the signal is represented by the color's intensity. The greatest value possible is dark red. From there, the amplitude value will decrease until it reaches the lowest level in the purple area.

$$\begin{bmatrix} a_{00} & \cdots & a_{0T} \\ \vdots & \ddots & \vdots \\ a_{F0} & \cdots & a_{FT} \end{bmatrix} \tag{2.1}$$

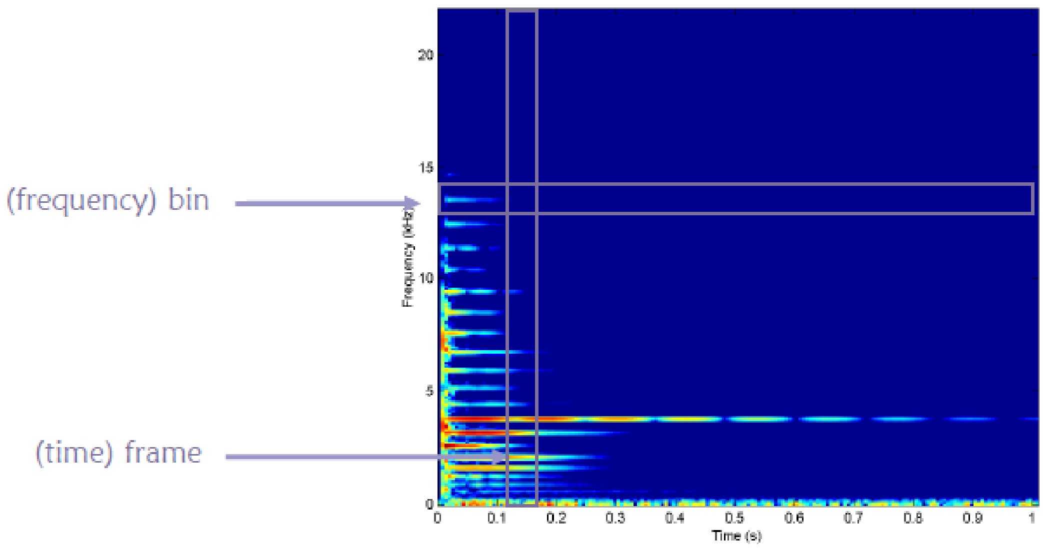


Figure 2.6 – From [Cavaco 09], a spectrogram.

3. Drum Kit and Cymbals

This chapter gives a brief overview of the different instruments that are included in the most typical drum kit setup, and of the different families of cymbals we intend to use in this work for analysis purposes. Since each drum and cymbal has its own characteristics and voice, it is of the utmost importance to cover their functions as an instrument in the drum set, and in the case of cymbals, the sound differences between them. This is the most important goal of this chapter; educate the reader in the sound differences between each class of cymbals, and how their very audible differences can actually translate into very hard characteristics for IFA to perform sound source separation accurately. This chapter will also serve as a very basic educational resource for those who would like to expand their knowledge on important sound features to consider when using feature based classification for cymbals.

We have also included a brief historical background on cymbals on the Attachments. This serves to show the importance of these instruments in different elements and eras of mankind's history, and how they evolved through time, helping to mold musical landscape from past and present alike. This further legitimizes the work developed for this dissertation, due to the level of historical and musical relevance of cymbals.

3.1. Drum Kit

The drum kit is considered as a collection of percussion instruments. In contemporary music more and more often we see all types of percussion instruments being mingled with the more usual western drum kit setup. When we talk about drum kits in this work, we only consider the most used and most common instruments found in the majority of drum kits - the snare drum, bass drum (played with the help of a pedal), hi-hat, tom-toms, crash and ride cymbals (figure 3.1). This setup is known as the rock/ pop drum kit.

Figure 3.1 – A basic rock/pop drum kit.

The snare drum, bass drum and hi-hat are the pieces that define the essence of a drum kit; they are the main instruments in almost all types of music. Jazz is an exception, since the ride cymbal has a more important role than that of the hi-hat. The remaining instruments are important as well but will depend mostly in the style of music played, and on the drummer's personal preference. The importance of these four pieces of the drum kit is due to them being mainly used to keep time during songs, playing beats and embellishments that complement the song. Since time keeping is the most important role of a drummer, these four instruments become essential. The toms are used more often for fills, which are rhythmic patterns played in between sections of songs (e.g., between verse and chorus). They prepare the listener and the band to the next section. They can also be used in beats, just like the snare, bass drum, hi-hat, and ride can be used in fills, but that is not their main functions.

3.2. Cymbals

The families of cymbals described next, have three different and unique striking zones (except for china cymbals in certain conditions, as we will see later), that enable the drummer to get three unique types of sounds from the cymbals. Those areas are the edge, bow, and bell (figure 3.2).

Figure 3.2 – The different zones to hit on a cymbal.

Each cymbal family's name is very recent. A catalog from 1948 of one of the most famous cymbal companies of our time, Zildjian, did not state their cymbals as being crash or ride, but distinguished them by their sizes (7 to 26 inches) and weights (Thin, Medium, and Heavy, just to name a few). In the next sections we will be taking a look at each class of cymbals that will be used for the analysis stages of our work. Here we introduce each class's origins, mains usages, and playing techniques. We will also get to discuss how their physical features forge the aspect of their respective spectrograms and sound. The next sections are based around the various chapters that can be found on [Pinksterboer 92].

3.2.1. Hi-Hat

The hi-hat is not a cymbal per se, but two cymbals that work together as one. One of them has its bottom side facing down against the bottom side of the second cymbal, which in turn is facing up. The two instruments are hanged on a hi-hat stand which has a pedal board (figure 3.3).

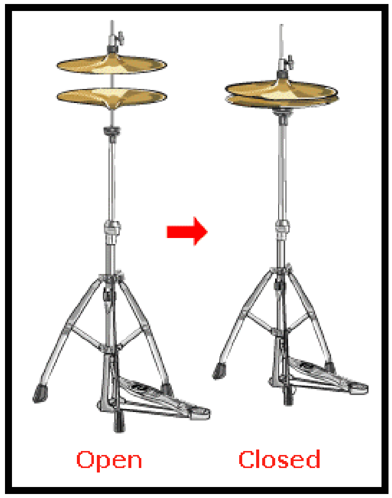


Figure 3.3 – A Hi-Hat.

The hi-hat is a very versatile instrument that enables the usage of a great number of techniques. When the pedal is pressed down the two cymbals are squashed against one another, this is the closed position or closed hi-hat (figure 3.3). When the pedal is not pressed down the two cymbals will have some distance separating them. This is called the opened position, or open hi-hat. The most common sizes for a pair of hi-hats range from 10 to 15 inches.

There are other techniques utilized with this cymbal like the “foot chick”; when the pedal is pressed down by the foot and a “chick” is heard as a result of the two cymbals hitting each other and closing the space between them; the “foot splash”, when the pedal is pressed and the two cymbals touch each other for a little fraction of time, returning promptly to the opened position.

In the next figure (3.4) we can see the spectrogram of a hit on the closed bow of the hi-hat, which resembles white-noise. The first thing you will notice in this spectrogram is that the energy level of this cymbal spreads along every value of the human frequency range with a very similar and fast decay. This behavior is very different from the one observed in the remaining cymbals, which have a longer decay that is not constant throughout the various frequencies. The quick white-noise effect we get with this cymbal is the result of the cymbals being closed when hit. As the drummer opens the cymbals the white noise effect continues, due to both cymbals rattling against each other with any stroke, but with a longer decay spread equally through the frequencies.

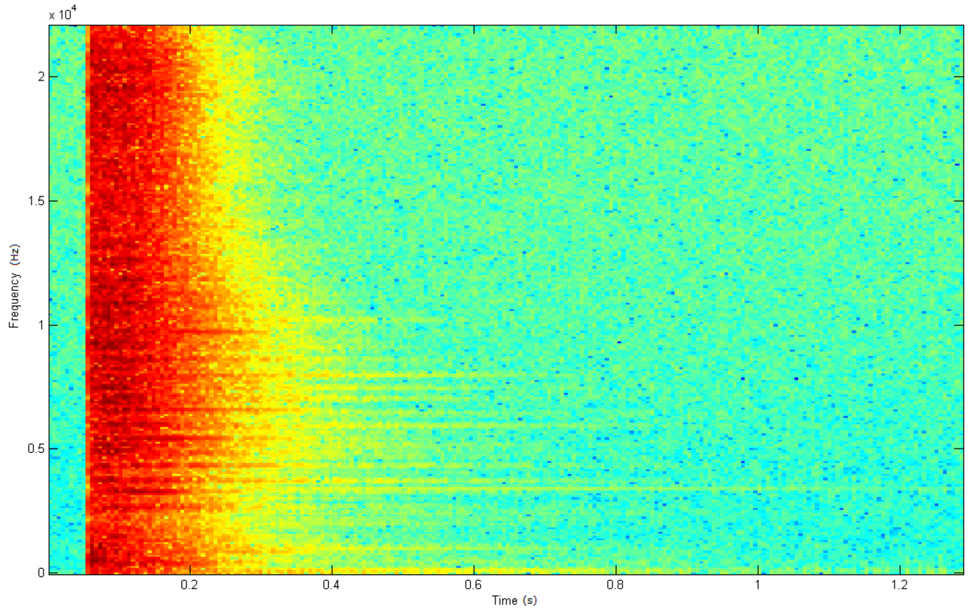


Figure 3.4 – Spectrogram of a hit on the bow of a Hi-Hat.

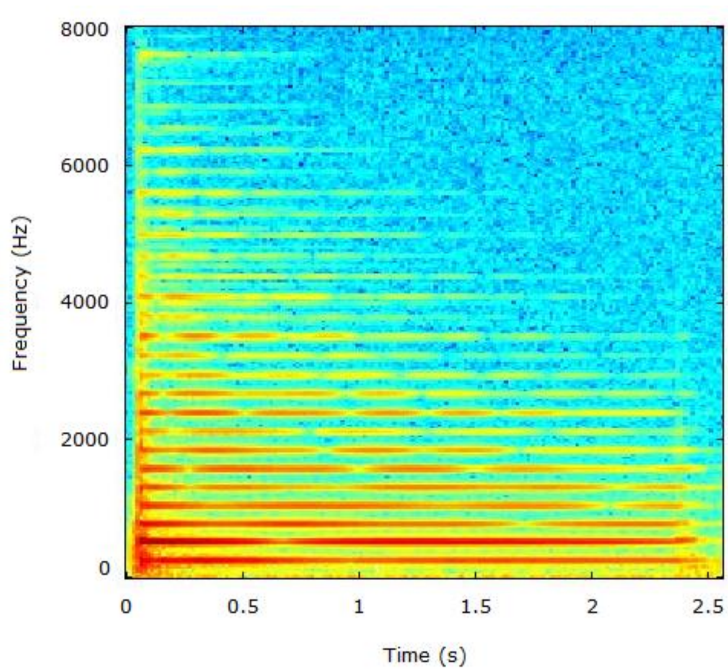


Figure 3.5 – Spectrogram of a note played on a piano.

Figure 3.5 is the spectrogram of a piano note. It gives us the fundamental frequency as the line with a higher level of energy, and the remaining harmonics from the note played. The difference between figures 3.4 and 3.5 is astonishing. In figure 3.5 instead of covering the entire human frequency range like on figure 3.4, we get very well defined bursts of energy. This is something common to any piano note. This way, it is harder to distinguish between the different cymbals than it is to distinguish between the different notes played on a piano.

3.2.2. Ride Cymbal

The name of this cymbal derives from what is played on it, steady, rhythmic, and driving patterns called ride patterns. That is why most drummers like to play this cymbal in the bow or bell areas, since these are the regions where we can get a more defined sound for playing the ride motifs. It is possible to find rides (figure 3.6) with sizes ranging from 18 to 24 inches. They are usually very heavy and thick, making their sound louder, compact, and much defined.

Figure 3.7 shows spectrograms of strong hits on both bell and bow areas of this cymbal. Taking a closer look at the spectrogram, we can see that the low frequency range (below 500 Hz) has a much longer decay. This is due to a couple of aspects - higher frequencies have a faster decay, low frequencies tend to last longer, and because all cymbals, when stricken, have an initial explosion that is rich in low frequencies. This does not mean the sound of this cymbal will be very low. However, due to their size and weight, ride cymbals tend to be lower pitched when compared with a crash cymbal, for instance, and as such have longer decays.

The differences between bow and bell can be observed on the spectrograms of figure 3.7. The bell sound is more compact, defined, and louder than the one from the bow. The amplitude levels on the spectrogram for the bell have more energy (they are in a very live red) than the same frequencies in the bow spectrogram (they are in a lively orange). As for the decay, there are various factors that determine the way it evolves in a cymbal. These factors are cymbal weight, cymbal size, bell size, and taper (change of thickness from the center of the cymbal to the edge) evenness[11]. All these factors contribute in one way or another for the overall decay of the cymbal. We would need a lot more information and study to be able really evaluate what is influencing the decay of both bell and bow. In comparison to the spectrogram of figure 3.4, these ride spectrograms are way more readable. They are still very noisy when compared with the one on figure 3.5.

Figure 3.6 – A Zildjian ZHT 20 inch Ride Cymbal.

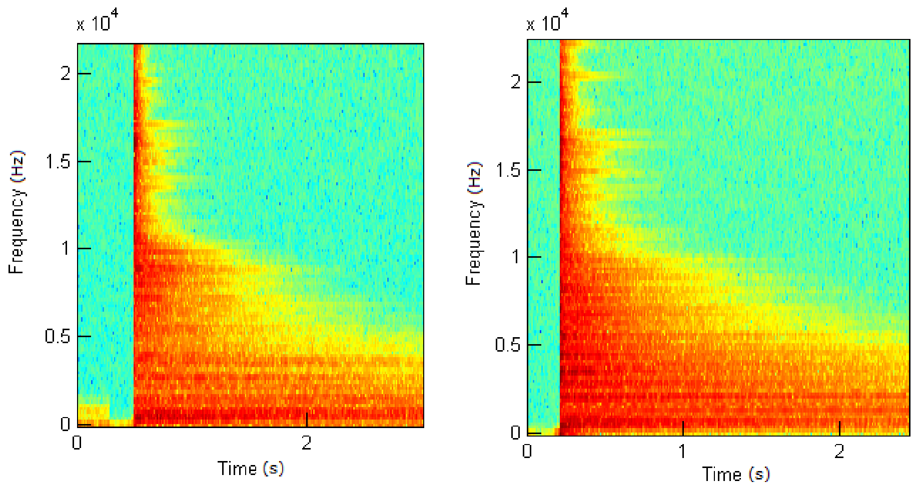


Figure 3.7 – (Left) Spectrogram of a hit in the bow of a ride.
(Right) Spectrogram of a hit in the bell of the ride.

[11] The decay increases with cymbal size, cymbal weight, the larger the bell, and with an even taper.

3.2.3. Crash Cymbal

After the development of the first ride cymbal, the smaller and lighter cymbals whose objective was of playing accents in a song by hitting their edges, eventually got named crash cymbals. These cymbals have a quick decay due to their usually thinner taper and lighter weight. The most common sizes for this type of cymbals are in the between 14 and 20 inches (figure 3.8), with the edge being the most played area of this type of cymbal.

Figure 3.8 – A Zildjian ZHT 14 inch Crash Cymbal.

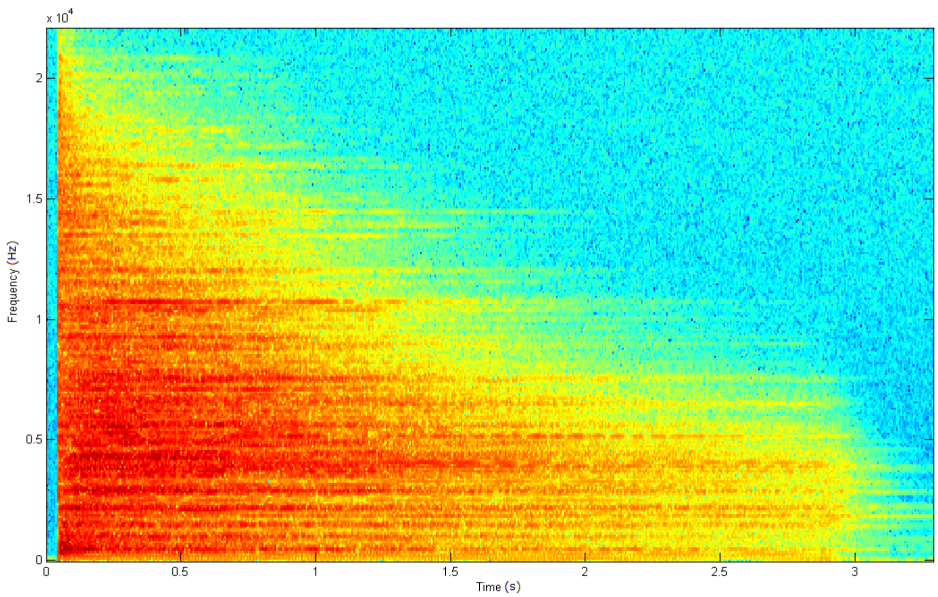


Figure 3.9 – Spectrogram of a hit on the edge of a crash cymbal.

Figure 3.9 shows the spectrogram of a crash cymbal when struck on the edge. When hitting this cymbal on the edge (known as crashing) the effect is a little different than when playing on the bow or edge of the ride. In the case of the crash, which is usually a much lighter and thinner cymbal than a ride, by striking its edge we will get more overtones, and a less controlled and defined sound. The decay is faster but the sound is explosive. Just like with the ride cymbal, the low frequency range has a much slower decay, with the higher frequencies having a faster decay, and low frequencies tending to last longer. However, there are a lot higher frequencies being excited here and with a longer decay than what we saw with the ride. Once again this is due to the weight and thickness of crash cymbals.

3.2.4. Splash Cymbal

These cymbals can be considered as small crash cymbals as can be seen on figure 3.10. Their sound is fast and bright, with a short sustain. Just like the crash cymbals, they are usually used for short accents. The most common sizes for splash cymbals are in between 6 and 12 inches.

Figure 3.10 – A Zildjian ZHT 10 inch Splash Cymbal.

The most used zone of this cymbal is the edge. Figure 3.11 shows the spectrogram of a hit on the edge of a splash cymbal. Both the higher and lower frequencies have very short sustain here, and even the explosion of the lower frequencies is mellower. This comes to show that as cymbals get smaller they tend to lose more and more of their lower frequencies. Thus their sound is predominantly high and fast, since the higher frequencies have a high decay.

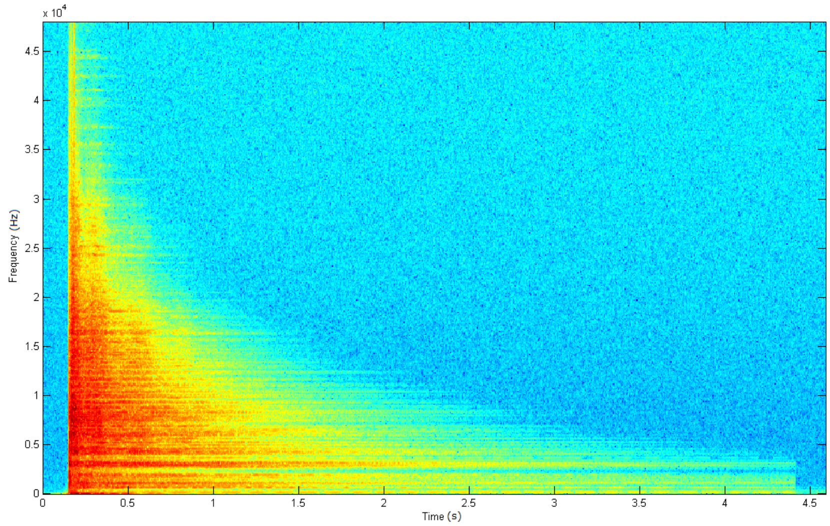
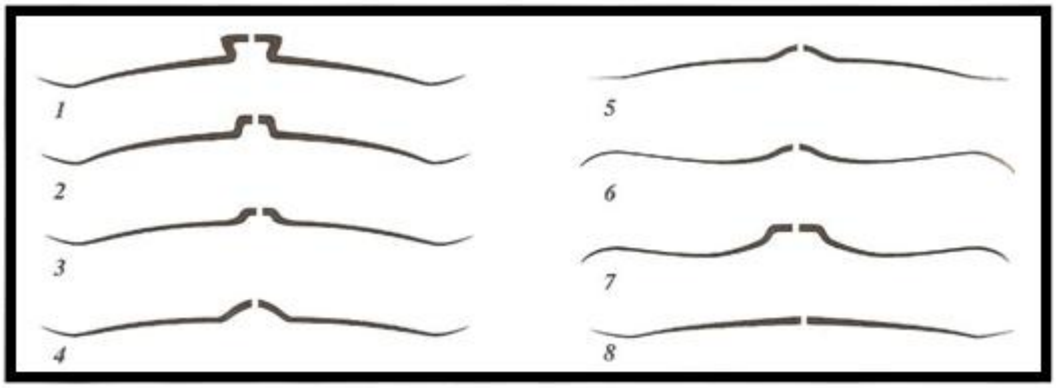


Figure 3.11 – Spectrogram of a hit in the edge of a Splash Cymbal.

3.2.5. China Cymbal

China cymbals where very popular at the beginning of the 20th century, and were used mainly as a ride cymbals. In the early 1970’s drummers started to use them more and more as additional crash cymbals. Like the name states, these cymbals came originally from China, and have a very characteristically flanged edge just like the cymbal of figure 3.12. The cymbal on the picture maintains very few resemblances with the original Chinese cymbals however, besides the flange.

Figure 3.12 – A Zildjian ZHT China Cymbal.



Approximate profiles of authentic Chinese Lion cymbals and some Western descendants
1. Chinese Lion cymbal (e.g. Wuhan) 2. China Type (Paiste) 3. China Boy (Zildjian)
4. Swish (Zildjian) / Chinese (Sabian) 5. Pang / Power Smash (Zildjian) / Flat Chinese (Sabian)
6. Novo China (Paiste) 7. Reversed China (UFIP) 8. Flat China (UFIP). Sketches by Ronald Willemsen

Figure 3.13 – [From Pinksterboer 92] Profiles of various types of china cymbals.

Original Chinese cymbals had a conical bell or handle, since these bells were used to be grabbed so a percussionist could crash a cymbal against each other. The western counterparts of the Chinese cymbals usually have a normal bell or a square one. Figure 3.13 shows the various shapes of china cymbals that can be found.

The sounds of some of the original Chinese cymbals resembled the sound produced by trash can lids. The western variations of this cymbal however are more pleasing to the ears, with a much warmer and harmonic sound. Nowadays these cymbals are most commonly used in the same manner as crash cymbal, but with an exotic sound to it; continuing a trend started in the seventies. Some drummers rather use it as a ride just like the first western drummers who used them. Due to its shape it can also be played in very different positions, whether facing up or flipped over. In this last position the bell of the cymbal cannot be played.

China cymbals have sizes that range from 6 to 27 inches. The sound of a china cymbal has a very fast decay, and just like the splash cymbal has a very bright sound, being however very piercing. Taking a look at the spectrogram below (figure 3.14) the initial amplitude values of most of the frequencies are in red, which comes to show how powerful the first moments of the sound of a china can be. The same rules we have been talking about with all the other cymbals apply here also.

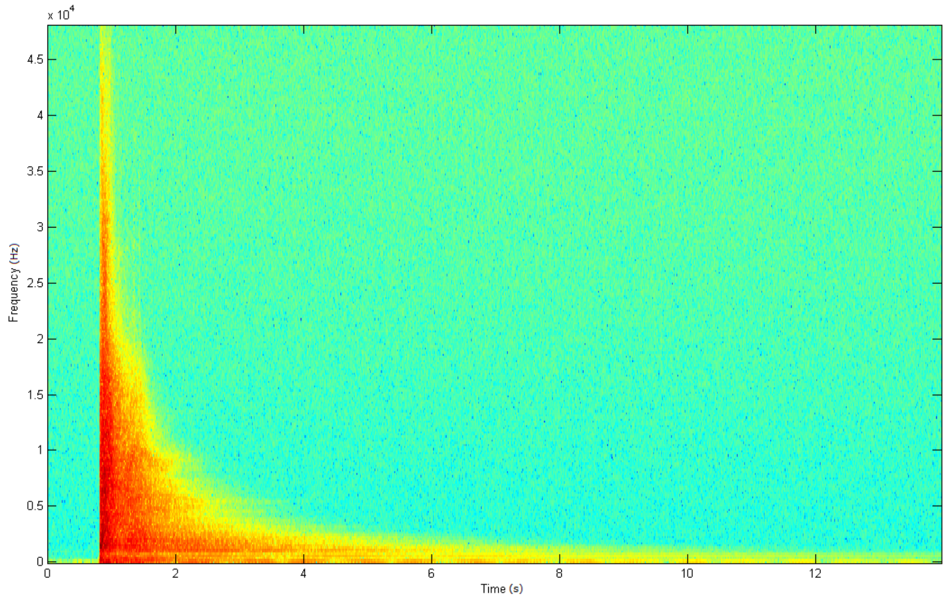


Figure 3.14 – Spectrogram of a hit on the edge of a China Cymbal.

4. State of the Art

Most classifiers studied for dealing with musical instruments are directed towards string and wind harmonic instruments. Still, some of these studies focus on the recognition of different types of strokes in percussion instruments with indefinite pitch, like the snare drum and conga drums [Bilmes 93][Schloss 85][Tindale 04]. However, most of the studies focus on identifying different instruments from the drum kit - bass drum, snare drum, hi-hat, toms and cymbals [FitzGerald 02][FitzGerald 04][Sillanp 02][Herrera 02][Gouyon 01][Paulus 06][Moreau 07]. Nonetheless, some of the proposed classifiers cannot clearly distinguish between the classes of cymbals. This means the sounds from any of the cymbals in the drum kit are assigned to the same class - cymbals.

Sound classifiers have two different stages, one for sound features extraction and another for classification. Many low and high level temporal, spectral and short-time features have been used to try to typify indefinite pitch percussion instruments. However, many classifiers give use to a blend of various features for getting good classification rates [Bilmes 93][Gouyon 01][Kaminskyj 01][Paulus 06][Schloss 85][Sillanp 02][Tindale 04]. This happens because of the issues that arise when deciding the most appropriate features to characterize the data. While most sound classifiers use a set of pre-defined features, others are that learn the features using decomposition methods such as ICA, ISA, Sub-band ISA, and NMF [FitzGerald 02][FitzGErald 04][Moreau 07], which we will be studying next, among another methods such as these.

4.1. Decomposition Methods

If I do not sit on a chair does it stop being a chair? If I use it as a table, will it be called a table from that moment on? What is it that makes a chair, a chair? Is it its shape or its function? After a while we realize that it is a very obvious answer - it is its shape, because even if we used a chair as a table for one hundred years, it would still be a chair being used as a table. But still, what is the principle that guides our assessment of reality that makes us decide that some object has a certain denomination?

When trying to figure out what defines a chair, we use inductive reasoning, i.e., an intellectual and conscious effort; however, to start this whole process of intellectualizing the chair, we have to first learn what a chair is. This is accomplished by perception [Attneave 54]. Perception is a sensorial mechanism that enables an inner representation of the outside world as well as its understanding. It enables us to react in the best possible way regarding external stimuli, having our own preservation as its main goal. Thus, speed on the perception of our surroundings is of the utmost importance. This can become a real problem to achieve, since we are constantly being bombarded with sensorial stimulus, and storing it all would be a total waste of space, since a great slice of our everyday stimulus is redundant, that is, accurately predictable and whose knowledge has already been acquired [Barlow 01]. But should the entire redundant stimulus be ignored to achieve a best level of comprehension about the new stimulus?

Barlow postulates that the perception of sensory messages may have a certain degree of redundancy and loss of information [Barlow 59], and that a total level of compression, that is, no redundancy whatsoever, is not the way our brain handles sensorial information. Without redundancy it would not be possible to identify structural regularities in the environment, essential to survival [Barlow 01]. This work developed by Barlow on quantification of information is called information theory. This discipline is instrumental in presenting compression techniques and redundancy reduction algorithms, not only useful in understanding how our brain functions, but in performing computer driven operations like image compression and sound source separation. A very well known case of sound source separation is described next.

In a cocktail party, the air surrounding our auditory sensors is cluttered with all the different confabulations taking place at one time. To this collection of sounds coming from different sources in the form of conversations, and engaging our ears as one single stream of cacophony, we call a signal mixture. Although some masking can occur, it is possible to concentrate on just one of those dialogues and separate it from the rest. This is known as the "cocktail party effect" [Arons 92] and is a problem of blind source separation (BSS). It is called BSS because there is an ability of separating a conversation from the mixture of dialogues without knowing the sources [Plumbley 02].

BSS is what we intend to perform in this work, but instead of separating one dialog from a stream of cacophony, we intend to identify to which class consecutively played cymbals in a signal mixture belong to. BSS based techniques use waveforms as inputs. Each one of the waveforms represents one source signal, and each source signal is a mixture of the sounds coming from the different sound sources. For each sound source there is a microphone recording the surrounding sounds. Now for our case, instead of using various waveforms we will use only one but represented by a spectrogram. A spectrogram can be assumed to be the result of the sum of an unknown number of independent source signals, each represented by an independent spectrogram. So in this chapter we take a look at some algorithms' potential to perform separation of sound sources form a spectrogram of a mixture of various cymbal samples.

FitzGerald made a very comprehensive study on the separation and classification of the standard rock/ pop drum kit's main instruments (check chapter 3.1 for more information on the rock/ pop drum kit). For that goal he used several algorithms, such as PCA, ICA, Independent Subspace Analysis (ISA), Sub-band ISA, and Prior Subspace Analysis (PSA), which we will explore in more detail below [FitzGerald 04]. Other promising techniques we will also explore include NMF and Non-Negative Sparse Coding, since they seem of great usefulness regarding cymbal separation.

We will start by analyzing PCA, ICA and NMF – that can be used as blind source separation algorithms, since they are what we like to call pure algorithms. This means they do not use other blind source separation techniques to achieve results, like ISA does. Afterwards, we analyze blind source separation techniques Sparse Coding, Non-Negative Sparse Coding, ISA, and Sub-band ISA. We will end this chapter with the analyzes of Locally Linear Embedding (LLE), an algorithm that can substitute PCA in techniques like ISA and Sub-band ISA, and with PSA.

4.1.1. Principal Component Analysis

PCA is a method used primarily for redundancy reduction or dimension reduction, i.e., data compression, and can be used to find patterns in high dimensional spaces. This is accomplished by finding an ordered set of uncorrelated Gaussian signals, such that each signal accounts for a decreasing proportion of the variability in the set of signal mixtures, where this variability is formalized as variance [Smith 02].

PCA starts by subtracting the mean from the N-dimensional mixtures in order to produce a data set with zero mean [Smith 02] (i.e., it centers the data at the origin of the N-dimensional space). Figure 4.1 illustrates this; on the left image we have the original N-dimensional mixture, while on the right we can check the result of subtracting the mean from the mixture. By not going along this line of procedure, the best fitting[12] plane will not pass through the data mean but instead through the origin [Miranda 07]. Once the data is centered PCA searches for the areas of greater variability, so that from a set of signal mixtures *x*, it can get a set of extracted/source signals *y*, that is, PCA tries to unmix the signal mixtures.

Lets us take as an example a 2-dimensional space, and two signal mixtures S'_1 and S'_2. From these mixtures it is possible to extract two source signals S_1 and S_2. For a successful extraction it is required to use an unmixing coefficient for each mixture. In this next example we use two of them, *a* and *b*, to extract S_1 like so:

$$S_1 = a \times S'_1 + b \times S'_2 \tag{4.1}$$

This pair of unmixing coefficients (a, b) defines a vector:

[12] Line/Plane of best fit, is a straight line/plane that best represents, or that best reconstructs (with minimum reconstruction error) the data of a given function/ scatter plot.

$$w_1 = \begin{pmatrix} a \\ b \end{pmatrix} \tag{4.2}$$

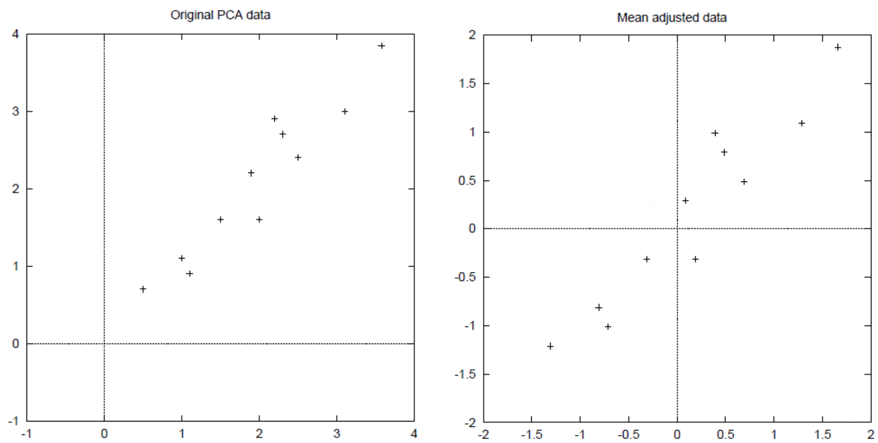


Figure 4.1 – From [Smith 02], Mean adjustment of the N-dimensional Space
On the left, the original mixtures on a 2-dimensional space.
On the right, the mean adjusted 2-dimensional space for the mixtures.

This vector has two very important geometric properties - length and orientation. Length defines the size of the amplitude of the extracted signal, making it bigger or smaller. Orientation is the factor that enables extraction of the signal. Let us call S to the space defined by the source signal axis S_1 and S_2, and by S' the space defined by the signal mixture axis S_1' and S_2' [Stone 04]. Both these spaces are defined in figure 4.2.

To unmix the signal mixtures we start by factorizing the mixtures by the employment of singular value decomposition (SVD). This technique decomposes a matrix into several component matrices that are often orthogonal or independent [Ientilucci 03]. The factorization goes like this, with *C* being the mixture matrix,

$$C = USV^T \tag{4.3}$$

U is a matrix with basis on the columns; S, a diagonal eigenvalue matrix; and V^T a matrix with time based source signals on the rows. The column vectors of U and line vectors of V are eigenvectors; with a related eigenvalue on the diagonal matrix S. Each of these vectors works just like the unmixing coefficient w_1, representing a line of best fit through the data mixture that finds uncorrelated Gaussian signals from it. Uncorrelation is assured by the orthogonality between the directions of the eigenvectors. Figure 4.3 has perpendicular vectors in red assuring uncorrelation, while the transformed axes are drawn as dotted lines.

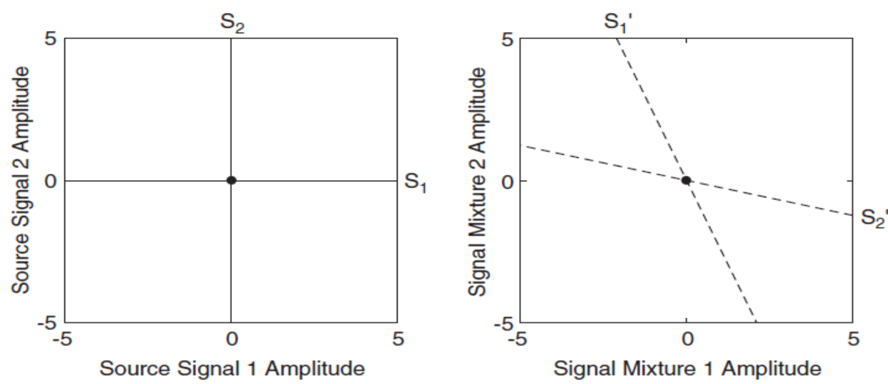


Figure 4.2 – From [Stone 04], source signal axis (left) and signal mixture axis (right).

Sorting the eigenvalues in descending order yields the same ordering for their respective eigenvectors on both U and V [FitzGerald 04]. This way, we will have the eigenvectors ordered from greatest to lowest value of variance [Smith 02]. This will enable us to perform data compression by removing the eigenvectors with the lowest values of variance, since lower variance dictates a less relevant eigenvector when it comes to the overall signal strength and idiosyncrasy.

Eigenvectors are scaled by the eigenvalues, this conveys that although their direction is untouched their size is not. This brings about one issue regarding not only PCA but ICA and NMF also; these algorithms do not accurately recover the amplitude information for each of the unmixed signals.

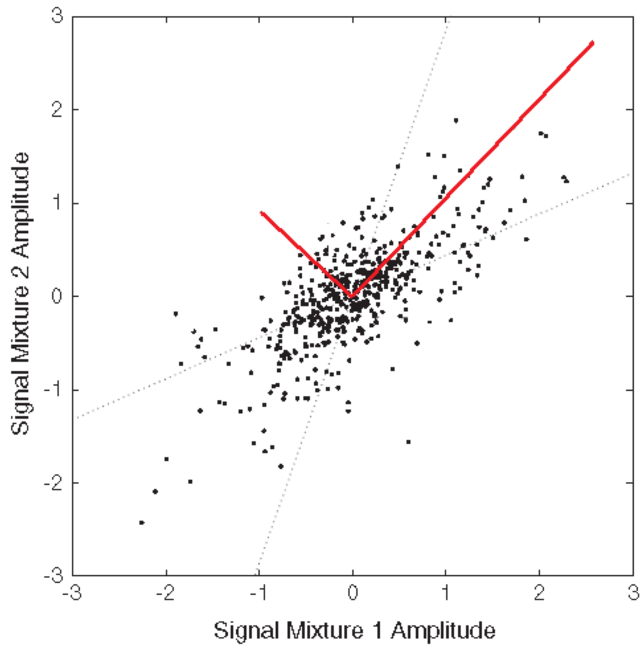


Figure 4.3 – From [Stone 04], PCA of two speech signals.
Each solid red line defines one eigenvector.

FitzGerald tested the use of PCA on spectrograms of drum sounds mixture. The information available on the spectrogram of the mixture is represented by a $(M \times N)$ matrix X with M signal mixtures. It is possible to learn a $(M \times M)$ unmixing matrix W that allows extracting M independent source signals from X:

$$Y = WX, \tag{4.4}$$

where Y is a $(M \times N)$ matrix that contains the M independent source signals. With $A = W^{-1}$, equation 4.4 can be rewritten as:

$$X = AY \tag{4.5}$$

where the columns of A are the basis that define the new space.

Figure 4.4 shows the spectrogram of the drum loop FitzGerald used. It contains sounds from snare drum, kick drum, and hi-hat. After performing PCA on the spectrogram we get a set of frequency basis functions. Figure 4.5 shows the first three basis functions, while on figure 4.6 we have the first three source signals. Each of the basis functions are related to any of the source signals; for instance, the first basis is related to the first source signal. This means that the source signals are the coefficients in a new dimensional space defined by the basis functions. The first frequency basis function is related to the whole signal, while the second and third show only information regarding the kick drum and snare drum sounds [FitzGerald 04].

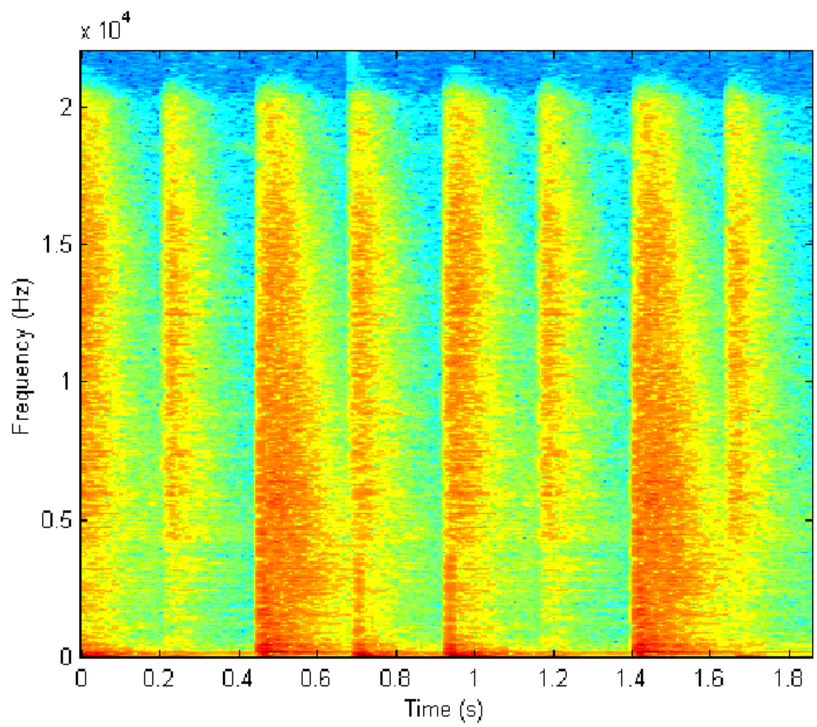


Figure 4.4 – From [FiztGerald 04], the spectrogram of a drum loop containing snare drum, kick drum and hi-hat.

We have a basis for snare drum and bass drum, but what about the hi-hat? This instrument has a very low amplitude level, so its variance is also low and the source signals that only have hi-hat information are ranked low. Clear information regarding hi-hats can be found only after five source signals [FitzGerald 04].

PCA may fail when performing individual sound source depiction due to it using orthogonal axes for separating the different sound sources from the mixture, something that may not be enough. There is no guaranty that it will separate the different sound sources in the mixture into separate source signals. This feature by itself is enough to discourage the use of PCA on cymbal separation.

The separation of each drum kit instrument through different basis was unsuccessful. This can be confirmed by the second and third basis functions and source signals of both figures 4.5 and 4.6, where information regarding both kick and snare is scattered through them. So even though PCA seems deemed to failure, there are ways of improving its overall success when separating the different sound sources from the mixture.

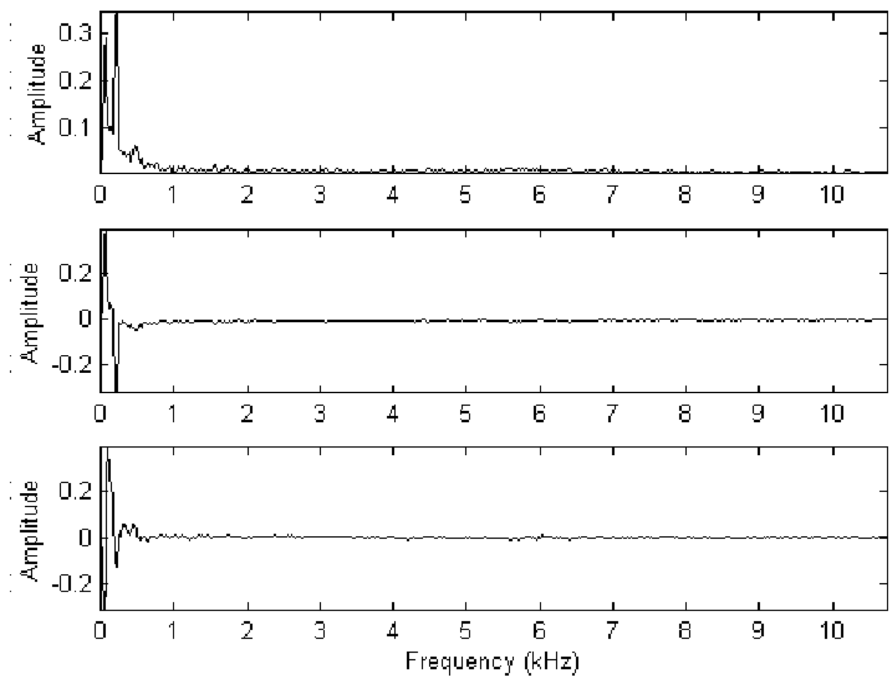


Figure 4.5 – From [FiztGerald 04], the first three basis functions.

Onset detection[13] could be used for the separation of each drum instrument through the search of abrupt increases in the energy envelope of the coefficients with the various basis [Hélen 05]. Afterwards the separated coefficients related to one specific drum kit piece could be joined in a single source signal. Anyhow there still remains a big problem, how to detach

13 Onset detection techniques detect the onset times of musical notes in audio signals. [Dixon 06]

coinciding events? Since this type of algorithm does not use prior knowledge but accumulated experience from the input, like we will see in NMF, if there are no isolated events that represent each of the drums in the coinciding event, separation is not possible [Smaragdis 03].

As we have seen, PCA favors basis of high amplitude. The information from sounds of low amplitude, like from the bow of the ride, or from a closed hi-hat can be represented by basis functions of very low rank.

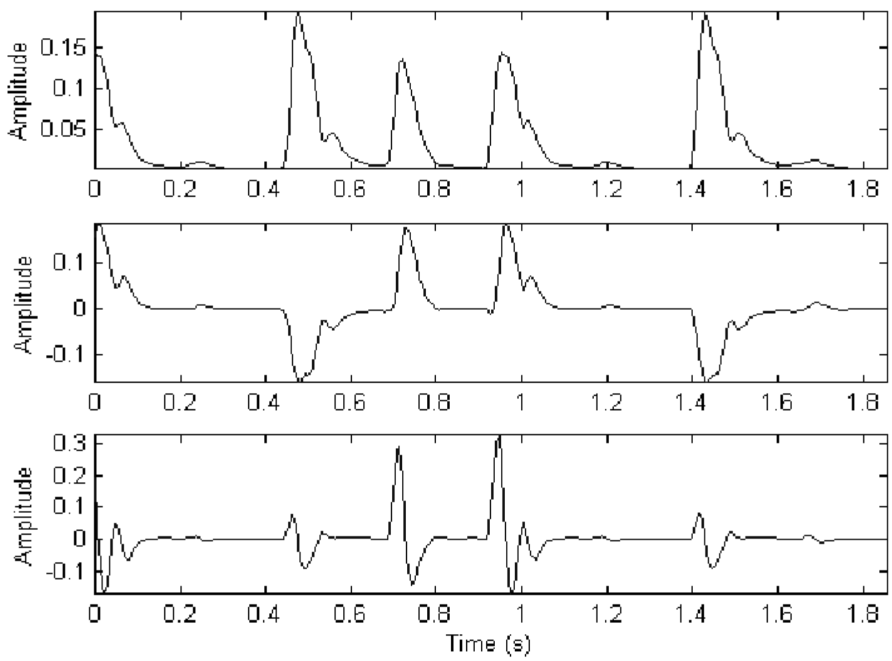


Figure 4.6 - From [FiztGerald 04], the first three source signals.

4.1.2. Independent Component Analysis

ICA can be used to identify the different sources in a mixture. While PCA tries to achieve this through the uncorrelation of source signals, ICA decomposes the signal mixture into a set of source signals through independence, a much stronger property than correlation [Stone 04].

When the mixtures are represented as waveforms, ICA requires having at least the same number of mixtures, that is, signals from different sound sources, as sound sources. For example, if we have two distinguishable sound sources, placing two microphones in two distinct places will create two different mixtures, since different distances of each sound source from the microphone will enable different proportions of the two signals in each mixture. Microphone placement works in the same way as camera placement. With an increased number of cameras filming a particular scene from different angles, we will get a much complete notion of what his going on. This way it will be possible to describe the scene with a greater level of detail [Stone 04]. However, when the sound of a drum kit is recorded in a studio[14] and ultimately mixed into a sound file, usually we get a maximum of two channels (stereo) from where we can separate the different cymbals used. Taking into account that we usually have at least three cymbals in a drum kit, ICA is doomed to failure if only two channels are available. To outflank this, another procedure can be used; much like PCA it is possible to apply ICA to the spectrogram of a sound mixture. Nevertheless with ICA the dimensionality of the data can be reduced by considering only I source signals, where $I < M$ [Cavaco 07].

To build the unmixing matrix it is required to use unmixing basis, one for each mixture. Thus, equations 4.1 and 4.2 are applicable here as well, and in the same molds, i.e., $w_{1,}$ which will be an unmixing basis in W, defines a weight vector used in the signal mixture space. Its length defines the size of the amplitude of the extracted signal, making it bigger or smaller. While the unmixed sound sources may be recovered, their original magnitude level can differ from the original signal. Orientation is the factor that enables extraction of the signal [Stone 04]. For a weight vector to extract a source signal it will have to be orthogonal to the orientations associated with the rest of the source signals, except the one that it will extract. In figure 4.7 we can see that by being orthogonal to S_2', w_1 will be able to separate source signal S_1, like stated.

[14] Check attachment "Drum Kit Sound Recording and Production" for more details on drum kit recording methods.

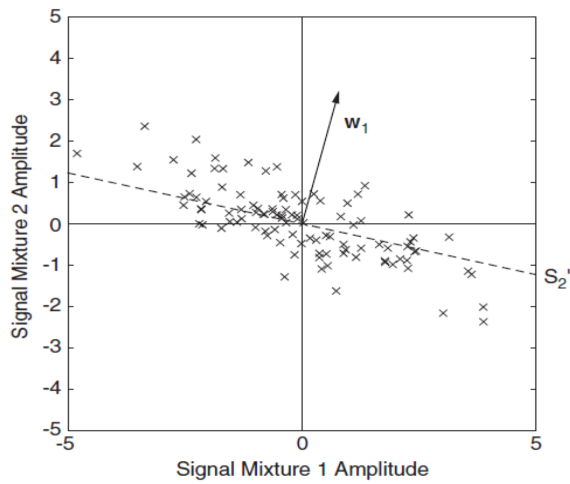


Figure 4.7 – From [Stone 04], w1 orthogonal to all source signals (S2) except S1.

method	SNR / dB
sinusoidal model	1.35
onset detection	3.05
ICA + SVM	-2.14
NMF + GMM	7.0
NMF + SVM (drums)	7.33
NMF + SVM (harmonic)	2.46
NMF + SVM (harmonic *)	7.33

Table 4.1 – From [Helen 05], SNR results for various types of sound source separation techniques.

Hélen performed the separation of an entire drum track from a polyphonic signal containing pitched instruments. The drum parts enclosed in the songs contained cymbals, tom-toms, snare, and bass drum. Hélen showed that it is possible to separate drum sounds from other instruments with both ICA and NMF of a spectrogram [Helen 05]. In addition, Hélen analyzed the level of quality of ICA's and NMF's separation using signal-to-noise ratio (SNR). With this type of measurement the level of the background noise is compared to the level of the ideal sound to unmix. The higher the values of SNR calculated the less influential the noise is over the signal, thus we have a greater level of success on the separation. The SNR obtained with all the methods are low, with ICA having the lowest value of them all, as can be seen on table 4.1. Other techniques like NMF, and under the same conditions, showed better performance than ICA when separating percussion instruments from the original mixture, in which cymbals were included.

4.1.3. Non-Negative Matrix Factorization

The base concept behind NMF is the same as the one seen on PCA and ICA. Nevertheless, rather than establishing statistical independence or uncorrelation as the basis for this factorization process, NMF uses non-negativity. This technique has a matrix notation similar to the one in equation 4.5, and can also be applied to the spectrogram of a mixture. Matrix X of size $(N \times M)$ is comprised of a set of N-dimensional data vectors, which are placed in its columns, with M signal mixtures in the rows. This matrix is then factorized into A of size $(N \times R)$ where its columns are the basis functions, and Y of size $(R \times M)$, with R source signals. This factorization is conceived in a way that makes it possible for the new matrices to be smaller than X, since $R \leq M \; and \; R \leq N$, which may result in data compression [Lee 01]. As we will see further down in this section, this can bring about some complications regarding the level of success of the factorization.

With the non-negative constraint. NMF does not allow negative values in any of the component's magnitude spectrums, enabling the components gains to be addictive between them. With this we have a parts-based representation, one that enables the different components to act like different parts of a source signal, without subtracting information between them to build the whole [Lee 99].

As an example of NMF application, Lee used this technique on a database of facial expressions as a way of learning how to represent a face as a linear combination of basis functions (entries in A) [Lee 99]. In figure 4.8 it is possible to witness in first hand NMF's effect over a picture of a face. X, A and Y are the same as the ones in equation 4.5. The reconstruction of the original image into X' shows the additive nature of this algorithm, and that with NMF the reconstruction of the original facial image loses its original magnitude values. This is shown by the levels of gray on figure 4.8, where these levels are different between the original image (X) and its reconstruction (X'). The original face is reconstructed accurately using the basis matrix, although being mostly an approximation of the original data.

A good example of using NMF for sound source separation comes from [Smaragdis 03]. Smaragdis and Brown performed a study on the transcription of a polyphonic music signal using NMF, where polyphony events were two notes played from one instrument at the same time and by the same instrument. This algorithm was tested over recordings of a piano, with both isolated and coinciding notes played. On figure 4.9 we can see a series of isolated notes and only one polyphonic event, which is surrounded by a red box.

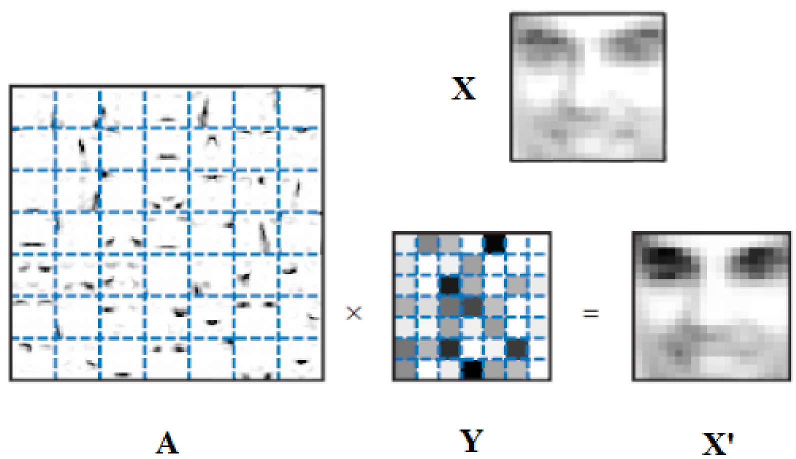


Figure 4.8 – From [Lee 99] NMF applied to face representation.

Figure 4.9 – From [Smaragdis 03], Musical piece played by a piano containing a polyphonic event with a red box surrounding it.

This musical piece has ten events with seven different notes, so let R be seven ($R = 7$). The result of NMF of this musical piece can be seen on figure 4.10. On the left image we have the representation of the values in matrix Y (source signals), and on the right the values in matrix A (basis functions). On the third row of Y we can observe a source signal filled with noise, which signals a non-note source signal. This non-note source signal is the result of setting R to seven, but having NMF consider that there are only six events. This means that one of the sources has two notes in it that are regarded as one event, instead of two. The notes we are talking about are the ones played at the same time in figure 4.9. You can locate them on the sixth row of Y of figure 4.10.

Since NMF does not use prior knowledge, the only way to achieve a comprehensive and correct transcription result is through accumulated experience from the input [Smaragdis 03]. Thus, for this technique to be able to separate those two notes in the mentioned sixth row, both of them have to be part of the musical piece as unique events also. Separation is not possible in this case since these two notes are always played at the same time in the input signal.

With this algorithm it is not possible to know exactly how many source signals are to be retrieved from the input signal without prior study of the musical piece. Setting a value for R will condition exactly how many source signals to be returned. If the value chosen is less than the number of notes in the input then information will be lost and exact reconstruction will not be possible. On the other hand, if R is greater than the number of notes available, the coefficients (notes) with greater level of energy can be distributed amongst the rest of the entries in A and Y. Ergo, the choice of a random value for R is not quite effective unless we know how many sources we want to retrieve from the input.

Moreau developed a system that presented a solution for the transcription of drum events using NMF. The events consisted of bass drum, snare drum, and hi-hat sounds. Table 4.2 shows the results of Moreau's efforts. Precision rate (R_p) is the ratio between the number of correct detections and the total number of detections; recall rate (R_r) is the ratio between the number of correct detections and the number of events in the reference annotation. The overall hit rate (R_h) was calculated as the mean of individual instrument hit rates [Moreau 07]. Probably the most noticeable aspects of this table are the results regarding the hi-hat, which are the worst from the bunch. The overall results were very poor, probably due to the test data utilized, since only a song of one minute long was used to test the system [Moreau 07].

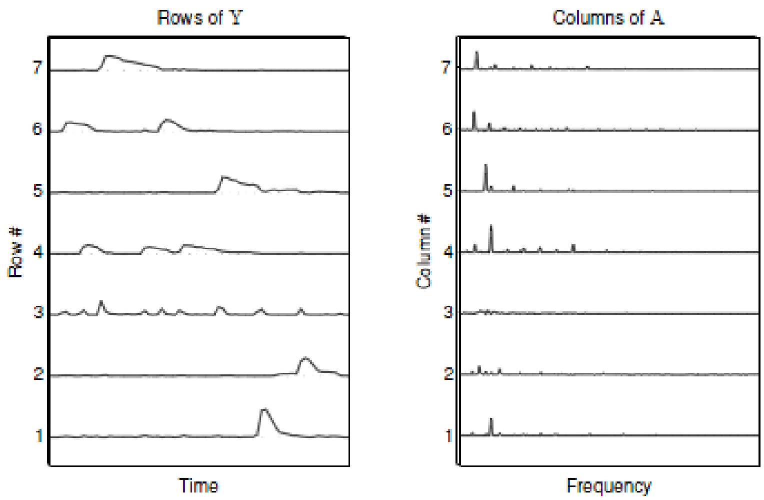


Figure 4.10 – From [Smaragdis 03], Decomposition of a musical piece.

NMF capacities were also tested along a system designed for the separation of a polyphonic musical signal into two classes - drum kit and pitched instruments [Hélen 05]. To achieve this goal the input signal was first separated into source signals using NMF. Afterwards support vector machines (SVM) classified sources according to one of the classes they belong to – harmonic instruments or drums. Results were evaluated using SNR. In the signals created for the testing phase, besides the usual drum kit pieces, bass drum, snare drum, and hi-hat, cymbals and toms were also added. The results can be seen in table 4.3. In this table it is possible to notice that from the algorithms tested, NMF with SMV gave the best results on the separation of the input signal into the two different classes. The results of the separation were not high with any of the methods, but NMF was the one that showed the greatest level of success. However, correct classification with SVMs of sources signals separated with NMF gave very encouraging results, with an accuracy of 93% [Helen 05].

	BD	SD	HH	mean
$R_p\%$	89.47	36.71	34.00	53.39
$R_r\%$	86.52	43.64	12.93	47.69
$R_h\%$	75.28	-47.27	-15.52	4.16

Table 4.2 – From [Moreau 07], decomposition results.
Rp – Precision Rate/ Rr - Recall Rate/ Rh – Instrument Hit Rate

method	SNR / dB
sinusoidal model	1.35
onset detection	3.05
ICA + SVM	-2.14
NMF + GMM	7.0
NMF + SVM (drums)	7.33
NMF + SVM (harmonic)	2.46
NMF + SVM (harmonic *)	7.33

Table 4.3 – From [Helen 05], SNR results for various types of sound source separation techniques.
In red the results of applying NMF of separating the drum part.

The last case studied was presented by Paulus and Virtanen [Paulus 05]. It consists of three stages. In the first one, source signals are estimated from training material for each instrument in the mixture. The training material comprises samples for unique sounds of each cymbal. NMF is applied to each sample for any instrument. The basis functions for samples pertaining to a given instrument are then averaged over the total number of samples for that cymbal in the training set hailing the instrument's source spectra. This procedure is repeated for all instruments [Paulus 05]. In the second stage each drum instrument is separated from the mixture using the training source spectra. In the last stage of the algorithm, onset detection is applied to determine the temporal locations of sound events from the separated Signals [Paulus 05]. As usual, snare drum, kick drum, and hi-hat were used for this test. In table 4.4, precision rate (R_p), is the ratio of correct detections to all detections; recall rate (R_r) is the ratio of correct detections to number of events in the reference annotation. The overall hit rate (R_h) was calculated as the mean of individual instrument hit rates. Avg is the result of adding the percentages of each instrument, regarding a type of rate and dividing it by the number of instruments, B (bass drum), S (snare drum) and H (hi-hat). NMF presents better results than PSA[15], especially on the hi-hat. So this algorithm may perform very well against cymbals.

unprocessed		B	S	H	avg
SVM	R_p %	99	93	92	94
	R_r %	99	93	86	89
	R_h %	98	86	77	**87**
PSA	R_p %	91	77	80	82
	R_r %	95	91	70	78
	R_h %	86	70	46	**67**
NSF	R_p %	100	100	98	99
	R_r%	100	94	96	96
	R_h %	100	93	94	**96**

processed		B	S	H	avg
SVM	R_p %	99	100	95	97
	R_r %	99	93	91	93
	R_h %	98	93	86	**92**
PSA	R_p %	77	83	80	80
	R_r %	92	84	73	78
	R_h %	71	67	51	**63**
NSF	R_p %	98	100	96	97
	R_r %	98	94	96	96
	R_h %	95	94	93	**94**

Table 4.4 – From [Paulus 05], table were PSA and NSF (Non-negative spectrogram factorization – NMF applied to a spectrogram) are applied on an unprocessed signal (left) table were PSA and NSF are applied on a processed signal (right).

The results of the analysis in [Moreau 07] although substantially weak, especially with the hi-hat, are insufficient to reach a conclusion, since only one test signal was used. On the other hand, the methods used in [Smaragdis 03] were successful in their separation efforts. Nonetheless they were not able to separate notes played at the same time. The only way to achieve separation with NMF is if both notes are part of the musical piece as unique events also. The notes played at the same time are one event, and is with events that NMF works. Like Smaragdis, Helén and Virtanen in [Helén 05] had a certain degree of success in proving that NMF could be effective in separating drum signals from polyphonic signals in a way that helped the classifier hail very good results, with a success rate of 93%. What is most encouraging is that besides considering the usual drum kit pieces for separation, cymbals were also added to the mix. The results in [Paulus 05] are very encouraging. AS you will able to see in section 4.8 of this chapter, PSA has very good results in what concerns separating bass drum, snare drum, and hi-hat from a mixture. However with NMF the results are even better, and the hi-hat, which is the cymbal that could be the most neglected here,

[15] Check section 4.8 for details on this algorithm.

actually has a success rate of 98% for unprocessed signals, and of 96% for processed signals, which is quite astonishing.

With the results shown here it is possible to admit that NMF may be a suitable algorithm to perform cymbal separation with some level of success. We don't have cymbals samples being played at the same time (they are played sequentially in the same sound file), so the issues found on [Smaragdis 03] may not occur. We also use a classification algorithm over the sound sources separated from NMF. Since in [Helén 05] we have a 93% of success when using a combination of NMF with a classification algorithm, and a 98%/ 96% of precision ratio for hi-hat detection, once again, from these results we expect this to be a very good option for classifying cymbals accurately from the mixture.

4.1.4. Sparse Coding and Non-Negative Sparse Coding

Sparse coding was intended to be a coding strategy that would be capable of simulating the receptive fields of the cells of the visual cortex of mammals [Olshausen 96]. Sparse coding considers that at a given moment only a certain number of sources are active, which means that only a certain number of sources are responsible for the creation of each observed signal [FitzGerald 04]. In order to identify the source signals sparse coding has to find the set of basis functions that enables the greatest level of independency amongst the source signals.

Olshausen conjectured that an image could be described with only a few coefficients out of the full set. To achieve this a form of low-entropy[16] should be found. If low-entropy is applied to all source signals, a lower level of dependencies can be achieved between them, enabling a greater level of sound source separation [Olshausen 96], and a greater level of independency. We first talked about independence when we introduced ICA for the first time, thus is there any kind of relationship between ICA and sparse coding?

16 Entropy is the level of uncertainty associated with a given variable. The higher the entropy, the higher the independence between the sources.

The model followed by sparse coding, is similar to the one already seen in ICA (equation 4.5) but with the addition of an error term (E) that accounts for noise (for instance in the signal transmission):

$$X = AY + E \tag{4.6}$$

This way, sparse coding does not try to recreate the original sources data perfectly, like ICA, focusing only on recreating it approximately [FitzGerald 04], with minimum reconstruction error [Olshausen 96].

The error term, a cost function[17], is the one responsible for the lowering of entropy on the coefficients of the source signals, enabling a greater level of independence between sound sources, and also performs a form of redundancy reduction [FitzGerald 04]:

$$E = -[preserved\ information] - \lambda[sparseness\ of\ a_i] , \tag{4.7}$$

where λ is a positive constant that levels the degree of significance of the second term - $sparseness\ of\ a_i$, relative to the first - $preserved\ information$. This term ($preserved\ information$) is the mean square of the error between the original and the reconstructed signal mixture, measuring how well the reconstructed signal describes the original mixture. The second term of equation 4.7 has a cost assigned to a_i[18] that depends on the level of activity that is scattered throughout the coefficients. Activity here is the level of participation of the coefficients in the reconstructed data. A higher cost goes out to a greater level of scattered activity. In the case of overlaps, this cost value forces the system to choose the coefficient most capable of describing a certain structure of the signal's data [Olshausen 96]. Sparse coding, like PCA, performs dimensional reduction, and may present problems with the separation of sound sources of lower level of amplitude [FitzGerald 04].

17 A cost function $E(q)$ is a function of q, which tells us what the minimum cost is for producing q units of output [Chan 07].

18 The group of source signals separated from the mixture, where each *i* is a source signal.

Abdallah and Plumbley tried to achieve automatic music transcription of an extract from a Bach piece played on a synthetic harpsichord[19] with sparse coding. The results were said to be passable [Abdallah 03]. Still we have to consider that the tests were done on a synthetic instrument with a very small data set, so it is yet to be seen how their system would behave with an acoustic instrument, and with a large data set.

Another test was made using non-negative sparse coding, that is, sparse coding where A, X, and Y of equation 4.6 all have non-negative column values [Virtanen 03]. But this time, instead of synthetic instruments, two acoustic instruments were selected: the snare and the bass drum. The transcription was tested using polyphonic signals containing pitched instruments synthesized from MIDI [FitzGerald 04]. This choice was made because through MIDI it was possible to have access to the correct drum score, not having to go through time consuming annotations to verify the final results obtained from transcription [Virtanen 03].

The transcription procedure starts by separating the most prominent coefficients. Then the identification of bass and snare sounds among the separated coefficients ensues, following the method described in the previous paragraph. Afterwards, onset detection is carried on the amplitude envelope of the source signals constructed from the coefficients, to detect the onset times of each hit on these two instruments. The performance of the transcription is evaluated using an error rate measure:

$$z = \frac{(N_d + N_i)}{(N_c + N_d + N_i)} \tag{4.7}$$

where N_c is the number of correct transcriptions, N_d is the number of deletions or missing events, and N_i is the number of insertions or extra events detected [FitzGerald 04].

Bass/ snare hits that are at most 32 milliseconds farther from the original hit are considered correct transcriptions. If a hit is determined as a snare or bass drum event, then they are counted as correct transcriptions. If however they are not recognized, but exist in the signal,

19 A musical instrument in which by pressing a key the chord is plucked instead of hammered, like in a piano.

they are considered deletions. If they are recognized but in reality are not part of the track, then are insertions. From the tests developed, there was an error rate of 27% for the bass drum and 43% for the snare drum [Virtanen 03].

As stated before, studies related to drum transcription are usually tested with a combination of bass drum, snare drum, and hi-hat. In this case the hi-hat was not used, because separation was very difficult due to their much weaker energy, compared with the bass and snare [Virtanen 03]. This predicament is the direct result of the redundancy reduction performed by sparse coding. Much like in PCA, cymbal separation may be very hard to perform with sparse coding, since knowing exactly how many source signals will represent the important information is a very big affair here. Therefore, when selecting the number of coefficients to maintain, information about elements with low amplitude levels may be lost. This is once again crucial to our intentions because in a mixture where we may have cymbals with low amplitude level, their information might be disregarded, and as such, separation is not possible which may difficulty the classification procedure.

In the analysis executed by Virtanen even the elements with high amplitude levels and of the same type, in this case skinned percussion instruments, were hard to separate with non-negative sparse coding. This way, it seems that when separating mixtures that have similar instruments, like skinned drums or in our case cymbals, the algorithm may have problems in separating the different sound sources from the mixture. Another problem arises from the lack of success of this algorithm for separating cymbals with low amplitude levels when stricken, like closed hi-hat and the bow of the ride.

4.1.5. Independent Subspace Analysis

ISA is a technique that was especially created to work with sound, in particular, it was developed to carry out sound source separation on a single channel apparatus. It first uses PCA to perform dimensional reduction on an input spectrogram and then ICA, so as to make the PCA source signals independent. The spectrogram is assumed to be the result of the sum of an unknown number of independent source signals, each represented by an (independent) spectrogram. These independent spectrograms are the result of the outer product[20] between a basis function and a source signal [FitzGerald 04].

Figure 4.11 shows the spectrogram of a sound clip containing a hi-hat, snare drum, and a piano. After applying ISA to the excerpt we get three source signals (figure 4.12), and three basis functions (figure 4.13). In each of the source signals of figure 4.12 it is possible to see that although separation was achieved, there still remains some unwanted information. The first source signal (snare drum) has some very small hi-hat peaks; the second source signal captures all of the piano notes, but we can see that the third one has some interference from the snare drum, since it coincides exactly with the snare stroke; the third source signal which is the hi-hat shows no problems. In each of the basis functions of figure 4.13 there is also unwanted information. In the snare drum, after the 1 kHz mark we have some residual noise, which in some part is related to the hi-hat. The second basis shows up the piano chord played as a set of peaks representing harmonics of the notes in the chord. The rest of this basis is a combination of noise with some characteristics from the hi-hat. The last basis has the main features of the hi-hat between 15 kHz and 20 kHz, with the lower frequencies of the basis having information regarding the piano [FitzGerald 04]. FitzGerald stated that after hearing the re-synthesis of the hi-hat, he noticed the presence of the attack portion of the piano notes, which is something that is missing in the re-synthesized piano signal. So, while the quality of the separation is good, overlaps between the separated source signals may happen, which to some degree may mask the separated signals.

Since ISA uses PCA and ICA to handle sound source separation, it is only natural that ISA inherits some of their limitations. In the dimensional reduction phase ISA neglects the source signals with a lower level of amplitude, which can make the recovery of sources like splash cymbals, rides played on the bow, and hi-hats a very hard task. This way, it may be necessary to increase the number of separated source signals, just to make sure that all the relevant information from different cymbals is maintained. This, of course, has repercussions in the robustness of ISA, since it is hard to set a correct threshold (number of components to

20 Outer product is the multiplication between two vectors, who's final result is a matrix.

maintain) since relative amplitudes of sources can vary from mixture to mixture, and even inside a same mixture, depending on the type of dynamics used by a drummer when playing.

The amount of information needed to perform sound source separation using ISA varies from signal to signal. This way, the number of dimensions to maintain from signal to signal in the PCA phase of the algorithm is unknown and will depend greatly on the amplitudes and frequencies of the sound sources [FitzGerald 02].

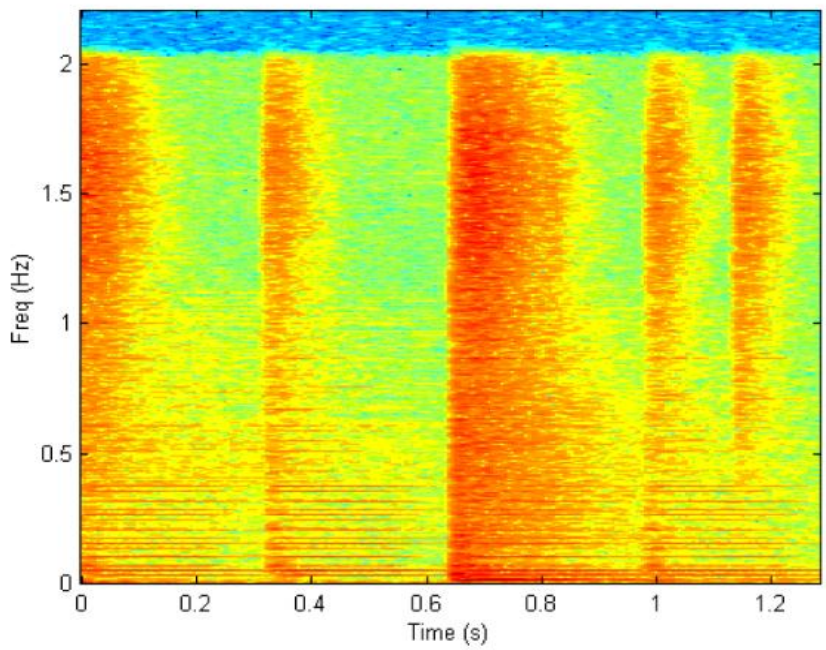


Figure 4.11 – From [FitzGerald 04], spectrogram of an audio excerpt taken from a commercially available CD.

In a signal containing only cymbal events, the usage of thresholds can be very risky since some cymbals may have much lower amplitude levels than others. Therefore, all cymbal coefficients related to a certain cymbal can be removed in the PCA stage of ISA, ending any chance of an accurate cymbal transcription – this is a limitation of PCA. Also, the coefficients that come from the ICA stage are not ordered in any way possible. This means that each of the coefficients has to be identified as being from a certain sound source, giving use to their frequency characteristics, or amplitude envelopes [FitzgGerald 04]. The big problem here is that, as we saw on chapter 3, cymbals show very similar frequency characteristics and envelopes, so even this identification of coefficients can go very wrong here. There are too many uncontrollable variables to attend to with ISA, which makes it seem like it is not the best choice for sound source separation of cymbals.

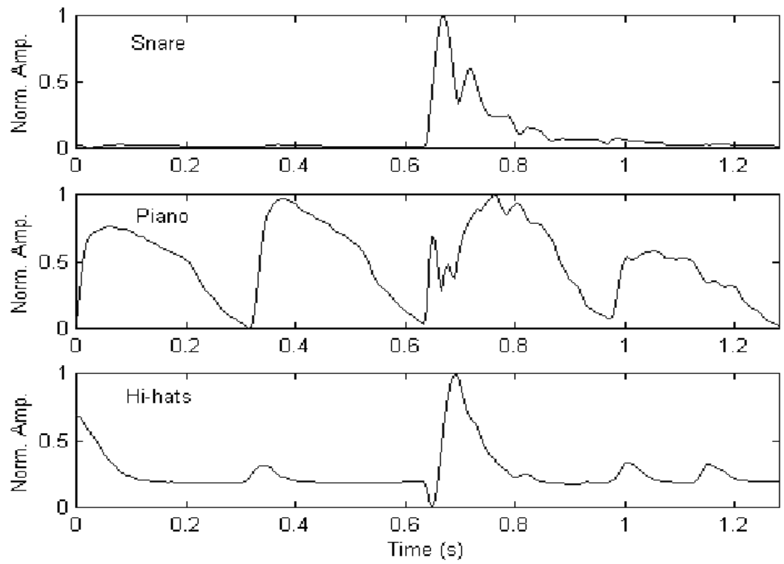


Figure 4.12 – From [FitzGerald 04], source signals for each of the instruments played on the signal from figure 4.11.

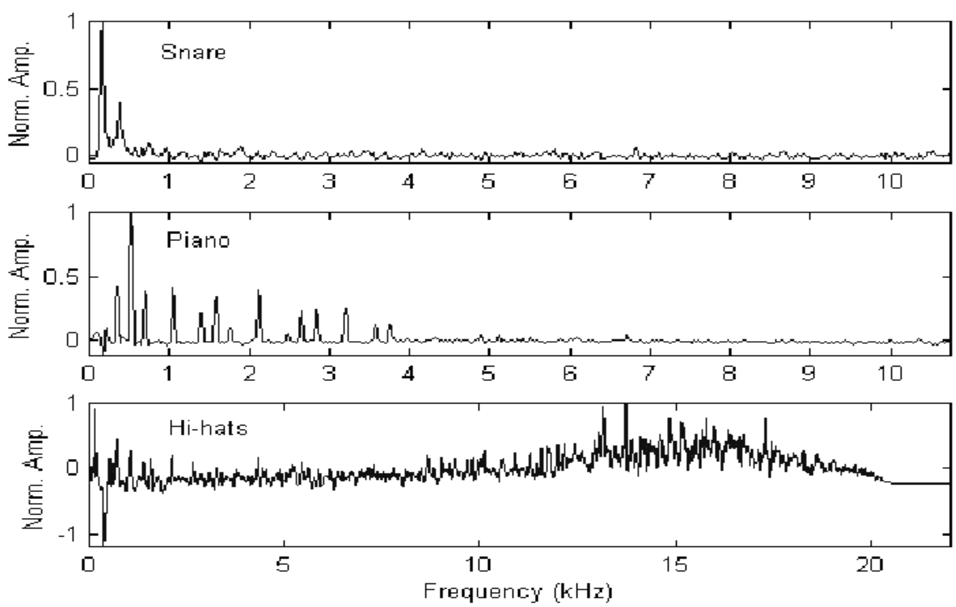


Figure 4.13 – From [FitzGerald 04], basis functions for each of the instruments played on the signal from figure 4.11.

4.1.6. Sub-Band Independent Subspace Analysis

This technique is based on ISA. The main difference consists of dividing the signal into sub-signals before performing ISA: the signal mixture is segregated in two sub-bands before performing ISA of each resulting sub-signal. FitzGerald performed tests with this algorithm on a drum loop with snare drum, closed hi-hat, and bass drum [FitzGerald 02 & FitzGerald 04]. The loop was severed into two sub-bands through one low pass filter with a cutoff frequency of 1 kHz, and a high pass filter with a cutoff frequency of 2 kHz, giving rise to two signals – one with a high frequency range and another with a low frequency range. This was the apparatus chosen because of the most important frequency bands that hi-hat (high frequencies), bass drum (low frequencies), and snare drum (low and high frequencies) cover. This may prevent the removal of cymbal coefficients from the overall signal, seeing they may become the events with a higher level of amplitude in the high frequency sub-band signal.

Applying sub-band ISA to the drum loop resulted in a whole collection of cleaner sound sources (with less noise). In addition the number of source signals required to recover the hi-hat was smaller than with ISA, as we will see next. Figure 4.14 exhibits the source signals retained by sub-band ISA of a drum loop, while figure 4.15 the coefficients from ISA of the same drum loop.

By comparing figures 4.14 and 4.15 we perceive that Sub-band ISA displays better results than ISA. With Sub-Band ISA the description of the three drum pieces utilized on the loop is done with less source signals, they are cleaner, and the hi-hat has more definition than with ISA. Despite its good results sub-band ISA is slower than ISA, since it requires two passes through the data, one for each sub-band. Since sub-band ISA is based on ISA one of the problems of ISA is still felt, which is the existence of more source signals than sound sources, but in a smaller number than with ISA [FitzGerald 02].

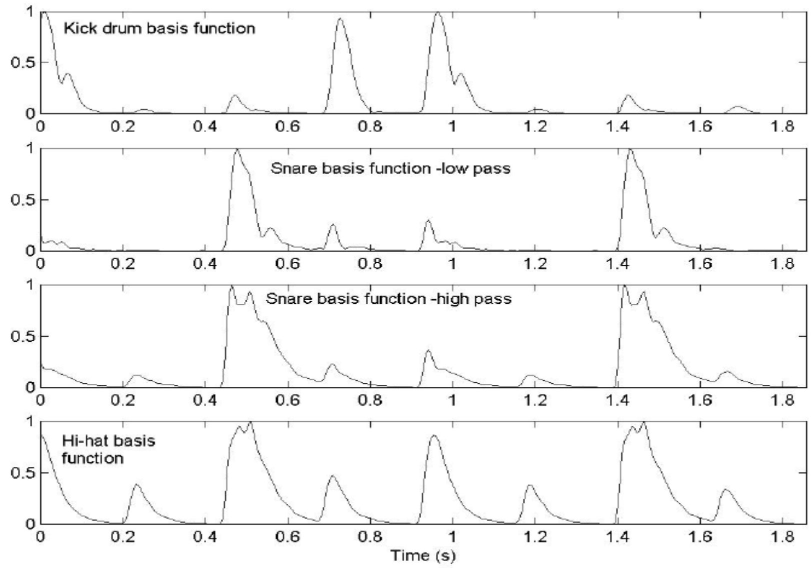


Figure 4.14 – From [FitzGerald 02], Sub-band ISA of a drum loop.

Type	Total	Undetected	Incorrect	%
Snare	21	0	2	90.5
Kick	33	0	0	100
Hats	79	6	6	84.8
Overall	133	6	8	89.5

Table 4.5 – From [FitzGerald 02], Sub-band ISA transcription results of a drum loop.

Table 4.5 exhibits the results of performing Sub-Band ISA on a drum loop. Total, refers to the number of total hits in each of the drum kit instruments present in the drum loop. Undetected, is the number of hits present in the sound mixture that were not detected. Incorrect, is the number of hits that were detected as being from the wrong instrument. Percentage refers to the percentage of accurate hits.

Although performing better than ISA and exhibiting very good results regarding the transcription of the drum kit events and even of the hi-hat, sub-band ISA still has a problem on the choice of the amount of information to maintain after the PCA phase of the technique, which would still be unknown. Consequently some important source signals might enable an accurate transcription could be lost forever. One other issue we found in this algorithm is that by applying it only to cymbals we would actually just be separating the cymbals frequency values through the two different sub-bands and not various cymbals for each sub-band. This is a result of cymbals having very “busy” frequency spectrums.

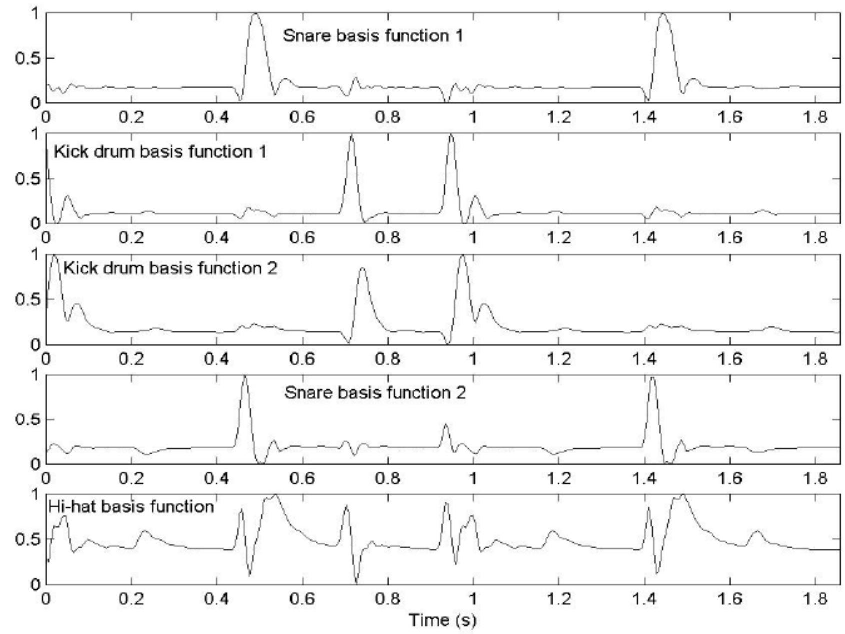


Figure 4.15 – From [FitzGerald 02], ISA of a drum loop.

4.1.7. Locally Linear Embedding

Locally Linear Embedding (LLE) can be used as a redundancy reduction technique, but contrary to other techniques studied in this chapter, it is not an information theoretic approach [FitzGerald 04]. This technique was included in this group of algorithms because of its possible applications in redundancy reduction, especially regarding its usage in ISA, where it can substitute PCA in the data redundancy reduction phase.

PCA's dimensional reduction is based around the concept of higher variance or higher amplitude level. As we have already studied, this may cost us the loss of important information related to sounds with low amplitude in the original musical piece. Important because this information may be related to cymbals. This loss happens because of the low level of power with which some cymbals are played, or because the area of the cymbal which is stricken has a natural low amplitude. When using ISA to perform sound source separation the first stage of the algorithm is performing PCA on the mixture, which may contribute to the loss of information from cymbals, something that is highly unwanted.

LLE is based on geometric principals, instead of the variance levels with which PCA reduces dimensional space. So, when used for dimensional reduction, LLE attempts to obtain a low dimensional space from the original high dimensional space, keeping the relative positions of data points, regarding its nearby neighborhood.

In a more mathematical approach, considering that the data is distributed to N real-valued vectors X_i with D dimensions, then we can consider that each vector and its respective neighborhood will lie on, or close, to pieces of data that can be characterized by coefficients that reconstruct each vector through its K-nearest neighbors (K-NN) [FitzGerald 03b]. To perform redundancy reduction with LLE, a value for the number of dimensions to keep (d) on the low dimensional space, will have to be specified, as well as the number of neighbors (K) to use for the reconstruction of each vector.

Because of the nature of this algorithm it can be combined with ISA to substitute PCA. To further test this assumption we take a look at a little test performed by FitzGerald on a drum sample containing snare drum, hi-hat, and bass drum [FitzGerald 03b]. In Figure 4.16 we can see the result of using LLE in ISA instead of PCA. The number of neighbors considered was thirty ($K = 30$) and the number of dimensions to recover from the signal was three ($d = 3$). The amplitude spikes match the correct locations for each stroke in each of the three drum kit pieces. When using PCA with ISA (figure 4.12) the results of the separation are well defined in the snare and bass drum, but however, LLE performs way better in separating the hi-hat.

ICA on the source signals of figure 4.16 results in an increase in the definition of each of the peaks (figure 4.18). The lower peaks on figure 4.17 may be due to the fact that the neighborhood points belong to other types of drums, or to drums with very similar frequency characteristics.

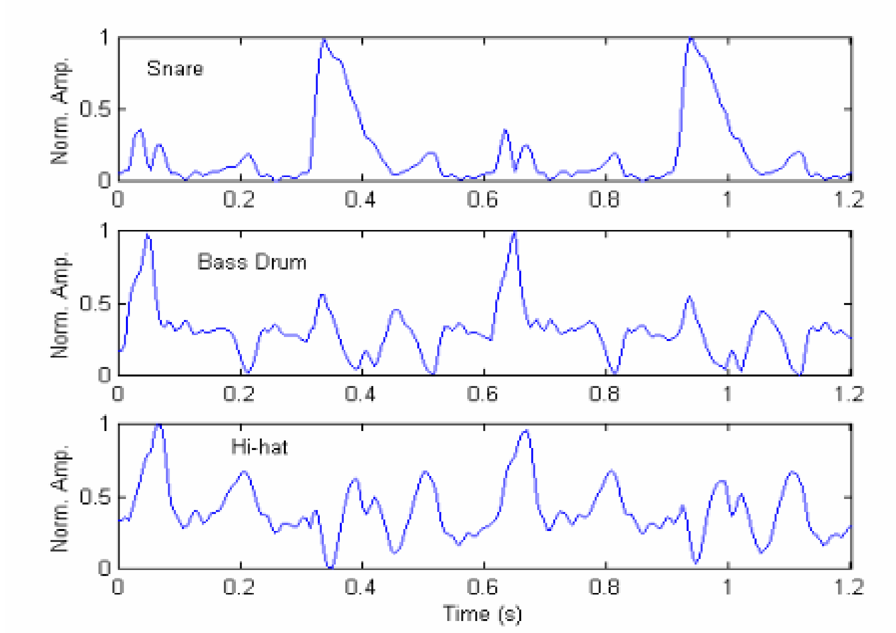


Figure 4.16 – From [FitzGerald 03b], source signals from using LLE in ISA instead of PCA, with K = 30 and d = 3.

As we have already stated, before performing LLE on the signal mixture we must first choose the number of neighbors to use in the reconstruction of the signal into a lower dimensional state. The choice of K when performing ICA on the output of LLE has to be done carefully, since, has we will be seeing next, the end results will vary with it.

Figure 4.19 shows the results of choosing a greater value for K than the one on figure 4.18. The third row of figure 4.19 shows that the hi-hat peaks are lower, while the ones that stand up the most are from the snare drum. This highlights that when using LLE in ISA much care must be taken when choosing a value for K, because this will influence the results of ICA. The problem here is that there is no way of choosing the most appropriate value for K, which would allow the technique to perform optimally. Nevertheless, FitzGerald stated that this problem is less harsh when the number of source signals recovered from LLE is greater than 10 ($d > 10$) [FitzGerald 03b].

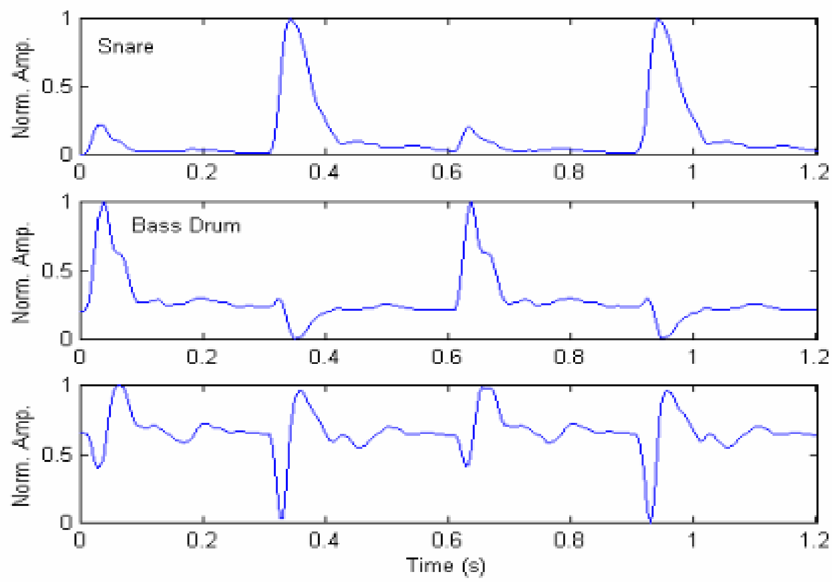


Figure 4.17 – From [FitzGerald 03b], source signals from using PCA in ISA.

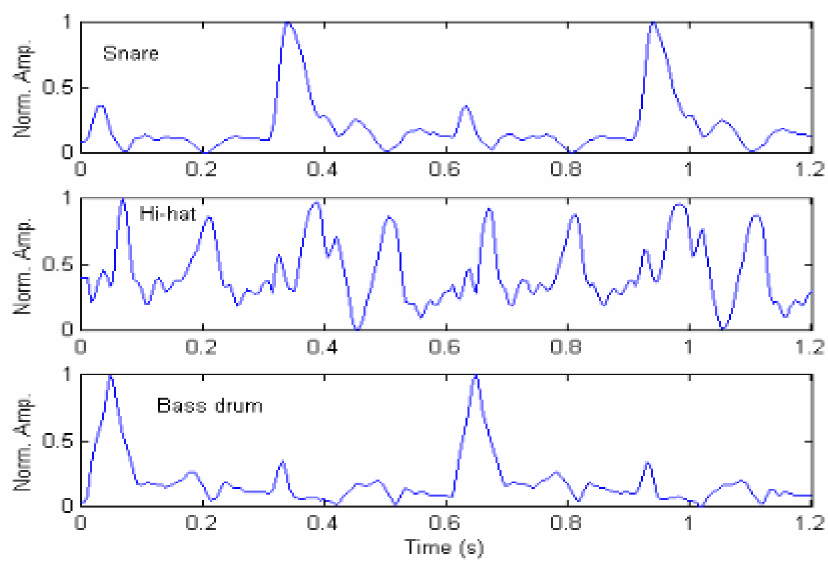


Figure 4.18 – From [FitzGerald 03b], coefficients obtained from ICA on the outputs of LLE, with K = 30.

First and foremost, the choice of values for K and d is done blindly as there is not a known value for K that can assure an optimal number of neighbors with which to reconstruct each vector. Anyhow, there is a way to bypass this situation if d is greater than ten, which in term, may create one other big problem. Since we do not know how many cymbals the input signals will exhibit, we could end up with a higher or smaller number of source signals than what is desirable for the separation. Furthermore, the number of source signals to output with LLE is likewise unknown, consequently the same problem that we had with ISA and sub-band ISA using PCA manifests in ISA using LLE, that is, not knowing how many source signals to input to the ICA phase of ISA. This will depend on the number of cymbals present in the mixture, something we are unaware of, since these algorithms are used without prior knowledge of what type of cymbals and how many are in the piece. Moreover, an additional problem abides in the neighbourhood of a given coefficient. As stated earlier, a neighbourhood may be comprised of a collection of data points, pertaining to different instruments, and whose frequencies spectrum superimposes one another in some values. When this happens the sources may not be characterised adequately [FitzGerald 03b], and since cymbals have overlapping frequencies, it may not be possible to guaranty a separation of cymbals through different coefficients.

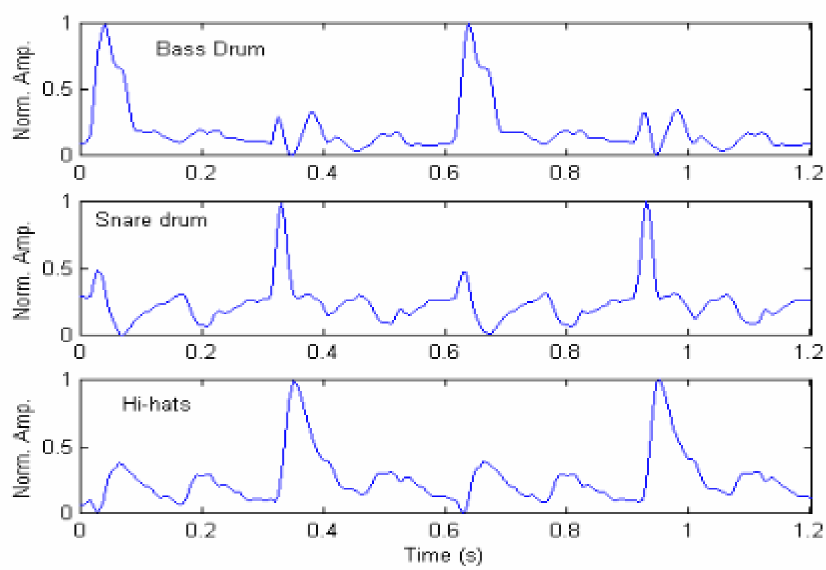


Figure 4.19 – From [FitzGerald 03b], coefficients obtained from ICA on the outputs of LLE with K =50.

4.1.8. Prior Subspace Analysis

PSA is the first technique in this work that incorporates models of the sounds in the mixture as training sets, as a way to achieve a better result in the separation of sound sources. The first step in PSA is to ensure the creation of a prior subspace capable of representing each sound source used in a given mixture. To do so, a large number of samples has to be analyzed for each of the instruments, in this case drum kit instruments, enabling the creation of a model for each instrument through a ISA type approach. This algorithm conditions each drum kit instrument to be pertained by a small number of invariants [FitzGerald 03a].

PSA starts by applying PCA to the spectrogram of each sample of a given instrument. The first three coefficients are then retained for further analysis. ICA is then applied to each one of the coefficients to get the independent frequency subspaces. This is so because the amplitude envelope of a pattern executed on a drum kit will depend exclusively on the way the drummer plays it, which varies greatly. The frequency values will be the ones chosen to represent the invariants of each drum, since this way we have a representation of a specific characteristic of the drum itself. The frequency subspace with the biggest variance is chosen to be the prior frequency subspace for that particular sample. After performing these operations on each sample for any instrument, K-means clustering is applied on the cluster of prior subspaces of samples for a given sound source. This way we get a prior subspace that characterizes each sound source.

After the prior subspace has been created for each of the drum kit's instruments in the source signal, their pseudo inverse are multiplied by the spectrogram of the input mixture. This originates the amplitude basis functions of each drum kit instrument in the mixture [FitzGerald 04]. Since drum sounds have a very noisy spectrum, each amplitude basis function may have smaller peaks from other instruments. To clean the functions from the unwanted peaks, and to get independent basis functions, ICA is used. This way, by multiplying the independent basis functions by its respective prior subspace we can estimate the independent spectrograms for each instrument in the mixture [FitzGerald 03a].

The use of prior subspaces naturally overcomes the problem of low amplitude sources, since PCA is not performed, unlike with ISA and sub-band ISA. This way, PSA has a faster performance level in comparison with ISA and sub-band ISA. Since sub-band ISA presented better results in comparison with ISA, we will further the correlation between PSA and sub-band ISA. To do so we will look at the tests made by FitzGerald for PSA, with the same drum loops used when performing tests with sub-band ISA [FitzGerald 04].

Type	Total	Undetected	Incorrect	%
Snare	21	0	2	90.5
Kick	33	0	0	100
Hats	79	6	6	84.8
Overall	133	6	8	89.5

Type	Total	Undetected	Incorrect	%Correct
Snare	21	0	2	90.5
Kick	33	0	0	100
Hats	79	2	6	89.9
Overall	133	6	8	92.5

Table 4.6 – From [FitzGerald 04], comparison between the results from applying sub-band ISA (left) and PSA (right) to the same drum loop.

As stated before, sub-band ISA performs ISA two times, one on the high-pass band and another on the low-pass band, while PSA only needs one pass performing in a much efficient manner. On table 4.6 we can see a comparison between the source signals separated by PSA and sub-band ISA. Total, refers to the number of total hits in each of the drum kit instruments present in the drum loop. Undetected, is the number of hits that although being in the sound mixture were not detected. Incorrect, is the number of hits that were detected as being of one instrument, when they belonged to another totally different. Percentage refers to the percentage of accurate hits.

With the use of a prior subspace, PSA is able to return a source signal for each sound source in the musical piece, outperforming sub-band ISA. PSA excels in the separation of hi-hat events, being 5% more successful than sub-band ISA (table 4.5). Even though the overall performance is better, there are snare events wrongly evaluated as hi-hat hits. This is due to a certain level of similarity between the higher frequency values of the snare and of the hi-hat [FitzGerald 04].

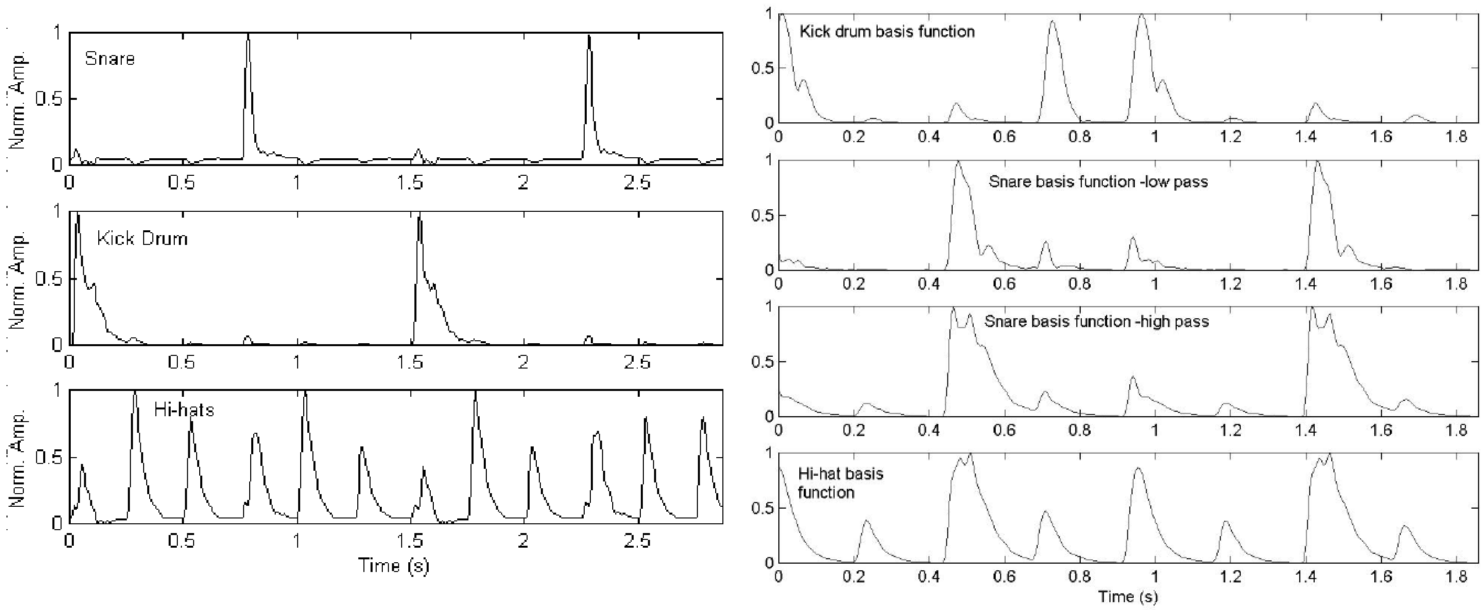


Figure 4.20 – From [FitzGerald 04], comparison between the source signals returned from applying sub-band ISA (right) and PSA (left).

In figure 4.20 the source signals that result from separation with PSA are cleaner, and are in the same number as that of the drum kit pieces. However, there may be some shortcomings when it comes to cymbals. The snare drum events wrongly evaluated as hi-hat hits, due to the level of similarity between the higher frequency values of both instruments, brings about an issue when it comes to separating cymbals. Since every cymbal has their energy spread along the human audible frequency range, it can become that much harder to separate the cymbals from each other, than to separate the snare from the hi-hat as shown on table 4.5. Either way, PSA seems to be a very good option for performing sound source separation of different cymbals, and a better one than ISA and Sub-band ISA.

5. The System

In the last chapter we reviewed a great number of algorithms. Each review was followed with a small analysis of their possible usage for detecting cymbal events. All of those analyses were just assumptions of what could be achieved by these techniques, since none of them had been previously used on the classification of cymbal events. Therefore, we have yet to see how they really work and behave in an environment filled exclusively with cymbal events. With that in mind, in this chapter we propose a system whose objective is performing sound source separation of the different cymbals in a signal mixture, and of accurately classifying them.

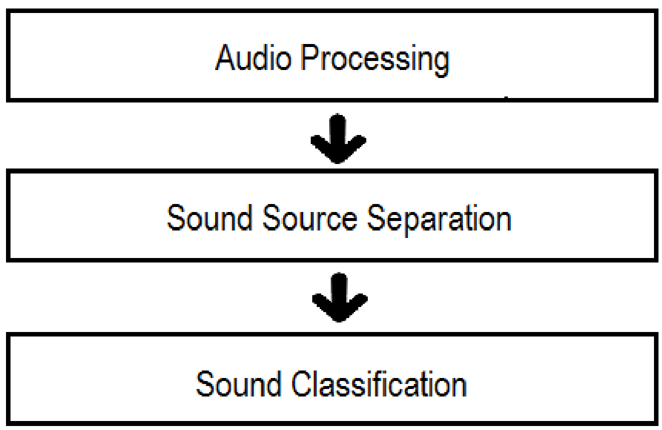


Figure 5.1 – Steps followed for automatic cymbal separation and classification.

The system follows an approach that consists of a three step sequence. The relationship between the three steps is displayed in figure 5.1. The audio processing stage comprises two different sub-divisions. The first one consists of converting samples from analog to digital (for further details on this conversion check chapters 2.1 and 2.2) through the recording of cymbal sounds into wav format. In the second sub-division the resulting waveform is transformed into a spectrogram. Then, in the sound source separation stage we apply one of the algorithms studied in chapter 4, to perform sound source separation. The final stage consists on using a classification algorithm, which categorizes each of the separated signals in the second stage into a cymbal class. In the following sections we analyze each one of these three steps in its own section with further detail.

5.1. Audio Processing Stage

For each analysis the system is fed with a combination of two or three classes of cymbals at the same time. Three possible combinations of samples were designed to be used for each one of the classes in the different cymbal combinations:

1. In the first sample combination, for each class of cymbals we choose the six samples with highest level of amplitude;
2. In the second sample combination, for each class of cymbals we choose the six samples which best describe the whole spectrum of amplitudes in a given cymbal, i.e., from high to low amplitudes;
3. In the third sample combination, for each class of cymbals we choose the six samples with the lowest amplitude level that still maintain sound characteristics particular to a peculiar family of cymbals[21].

For each analysis performed between cymbals, the sample combination chosen for a given class of cymbals has to be the same for the remaining classes. This means that if we use sample combination (1) for one cymbal class, then all the remaining cymbals will have to be tested with the same combination. This is how we organized the training set of our system.

After being transformed into magnitude spectrograms by applying STFT to each sound samples, the samples are concatenated as rows in a matrix. We do not consider phase

[21] When you are hitting a cymbal using very quiet strokes, it gets to a point where its sound does not emanate any audible characteristics associated with that given class of cymbals.

information here and use only the magnitude, taking the absolute values of the FFT spectra. All of this gives birth to the magnitude spectrogram matrix of the cymbal mixture.

5.2. Sound Source Separation Stage

Once the data is represented with a spectrogram we use NMF for performing sound source separation. As discussed in section 4.3. NMF presented encouraging results when it comes to sound source separation, which is a good indication that it can also mitt our goals. We know that the non-negative constraint is very useful in attaining the factorization of the whole, i.e., the mixture into its parts. Keeping that in mind, we followed a similar route to the one proposed by Virtanen [07], which in turn is based on Lee's and Seung's work [Lee 01] for using NMF for sound source separation.

NMF of the spectrogram of a mixture results into two non-negative matrices – A and Y. The product between these two matrices is equal to the spectrogram X, as in equation 4.5. All entries on A and Y are initialized with the absolute value of Gaussian noise. Estimation of both matrices is done by a cost function $c(A, Y)$, whose minimization algorithm tries to deprecate for each iteration of the factorization. This way, the reconstruction error of the product between A and Y vis-à-vis X is minimized. The cost function is a weighed sum of three terms – reconstruction error $c_T(A, Y)$, temporal continuity $c_t(Y)$, and sparseness $c_S(Y)$. Ergo, the cost function is,

$$c(A, Y) = c_T(A, Y) + \alpha c_t(Y) + \beta c_S(Y), \tag{5.1}$$

with α and β as weights for the last two terms [Virtanen 07].

Information theory algorithms are usually more sensitive to high-energies, failing to separate source signals with low-energy levels (PCA is a good example of this, as we have seen on chapter 4.1). Lee and Seung [Lee 01] tested two approaches for minimizing the reconstruction error of NMF – the square of the Euclidean distance and the divergence. The usage of a divergence is the best choice for our case, due to its sensitivity to low levels of energy when compared to the Euclidean distance [Virtanen 07]. So due to some of the cymbals used in our work, like the hi-hat, the ride, and the splash, this was the best option for us and the one we chose to use in this dissertation.

Adjacent time frames in a spectrogram show some continuity on their temporal structure, so they are not completely unrelated between themselves. The temporal continuity of the components is measured by assigning a cost to large changes between the gains in adjacent frames, which may be able to improve the separation between the source signals [Virtanen 07].

The last term of equation 5.1, sparseness, is taken in consideration because it helps to increase the weight of the redundant information, i.e., the most informative data, in the overall information landscape of a spectrogram. This way it may also increase the quality of sound source separation. To understand how this may happen let us look at a practical example using as subjects two instruments from the drum kit, the bass drum and the snare drum. Looking at the spectrograms of both instruments on figure 5.2 (bass drum on the left and snare drum on the right), it is possible to notice an overlap in the lowest portion of the frequencies from 0 to 1000 Hz. The overlap means that both instruments have energy along that same frequency interval. If we created a mixture with both these instruments, and used the sparseness criterion, the overlapped information of the bass drum would cover the lower frequency range of the snare drum. However, by giving use to sparse gains it is possible to model the snare drum with the information from the bass drum, plus the residual from the snare's higher frequencies [Virtanen 07].

While the ideas and possibilities behind the cost function are very interesting, as shown by Virtanen, the end result of its application can be far from the expected [Virtanen 07]. The objective of the work developed by Virtanen was to separate drum kit sound sources and pitched instruments sound sources from a mixture. For testing sound source separation using the apparatus we just described, of NMF and a cost function for minimizing reconstruction errors, Virtanen generated signal mixtures by using a random number of drum and pitched instruments sources. For the pitched instruments sources an arbitrary instrument and a fundamental frequency were chosen from the available samples, while for drum sources a random drum kit and a different drum instrument. Both the temporal continuity and sparseness terms of the cost function did not improve the results significantly.

Figure 5.3 shows the effect of applying different values to the weights α and β, when β and α are set to 0 respectively. In the y axis we have two different measures of success for the separation procedure; one for measuring the signal strength relative to background noise known as signal-to-noise ratio (SNR), and another for determining the degree of errors called error rate. There are three lines exhibited in the picture; the dashed one is for pitched sounds, the dotted line is for drum sounds, and the solid line represents the average results between the drum sounds and the pitched sounds. In regards to sparseness the figure shows that drum sounds have very low SNR (close to 3 dB) and very little error rate fluctuations. The only variation are when $\beta \geq 10^2$, where the results start to degrade, due to the size of the weight. With our case we are interested in the dotted line, since it was the one used for drums[22].

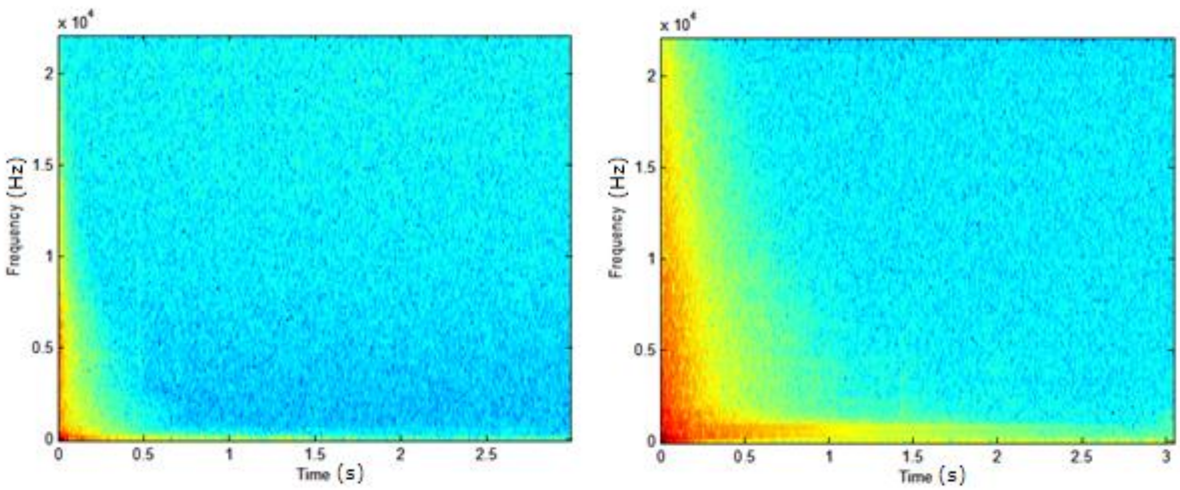


Figure 5.2 – Spectrograms of a stroke on a bass drum drum and on snare drum.
Bass drum spectrogram on the left and Snare drum spectrogram on the right.

[22] Samples from the sample based drum software synthesizer Drum Kit from Hell, developed by Toontrack, were used in [Virtanene 07], for both cymbals and drum sounds.

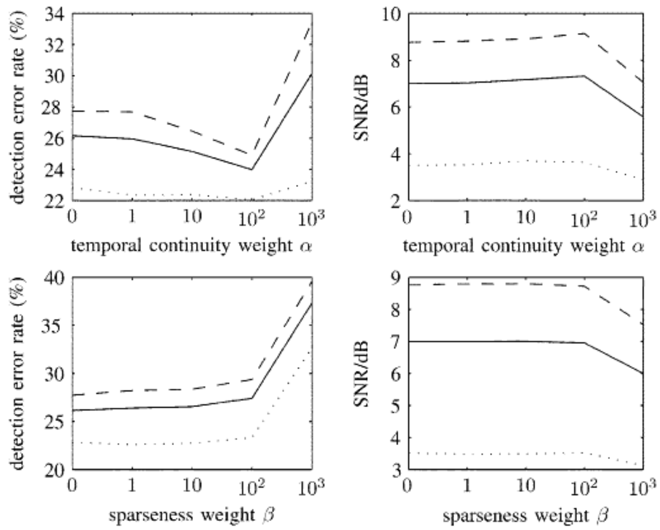


Figure 5.3 – From [Virtanen 07], effect of different temporal continuity weights α and sparseness weights β on the detection error rate and SNR.

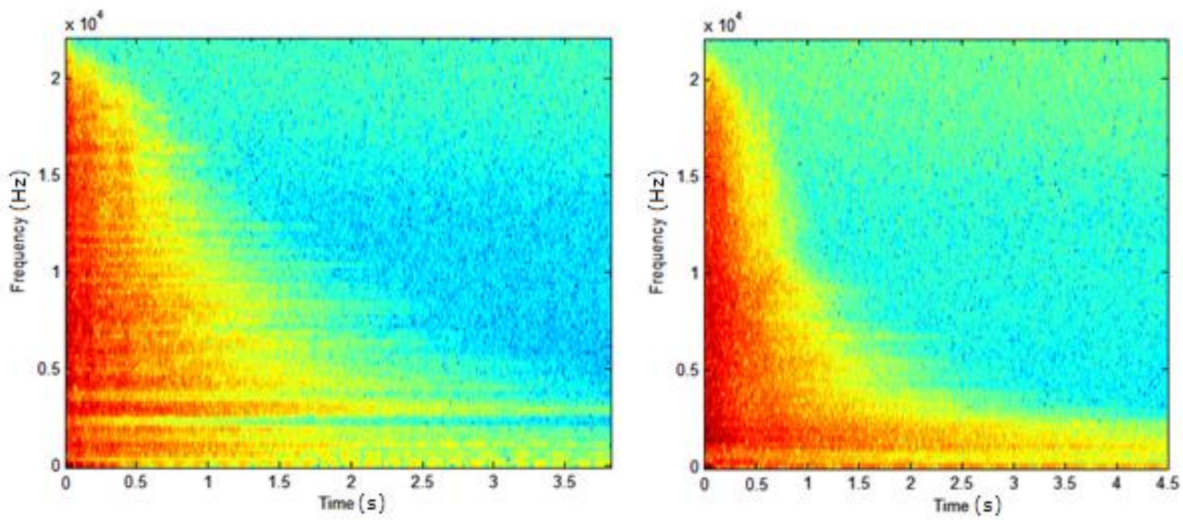


Figure 5.4 – Spectrograms of powerful strokes on the edge of a splash (left) and of a china cymbal (right).

The example given on figure 5.2 with bass and snare drum sounds was meant to illustrate the sparseness criteria as explained by Virtanen. However, if we take a look at the spectrograms of the two cymbals we can further understand how this sparseness criterion can actually fail. Figure 5.4 shows the spectrogram of a powerful hit on the edge of a splash cymbal (left), and a powerful stroke on the edge of a china cymbal (right). As can be observed there is much useful information in every bin and frame. Even when the samples of both china and splash have medium or low amplitude, like on figure 5.5, there is a lot of activity in both bins and frames.

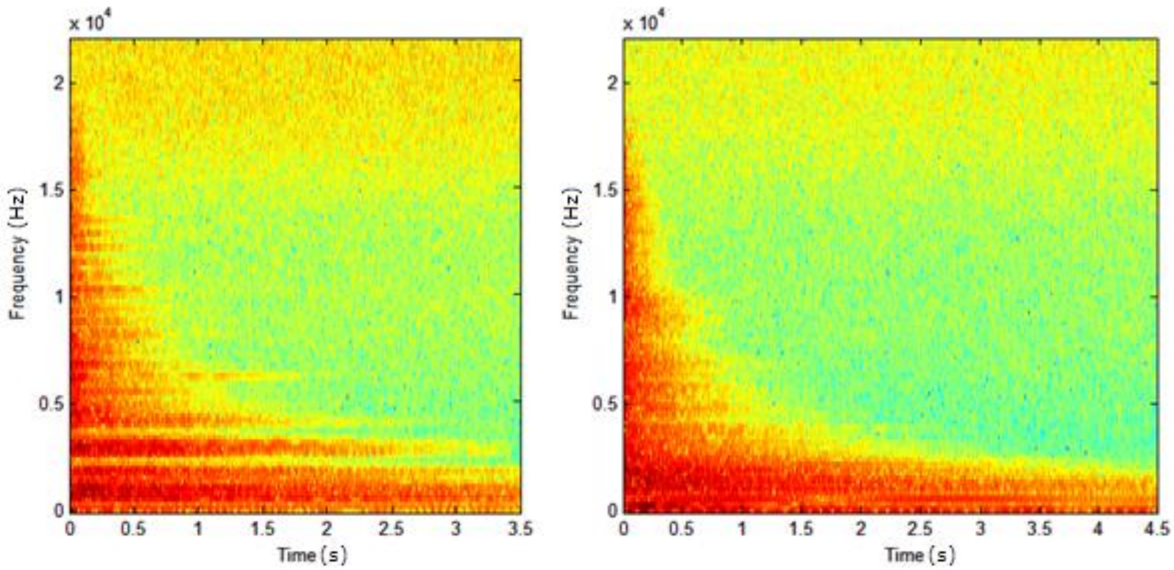


Figure 5.5 – Spectrograms of softer strokes on the edge of a splash (left) and of a china cymbal (right).

Temporal continuity, one of the terms of equation 5.1, shows that the results are almost identical to the sparseness results. We have done some preliminary tests that confirmed the insignificance of these terms – sparseness and temporal continuity. By increasing both α and β while keeping β and α equal to zero respectively, we found no differences in the success of the sound source separation results over the end result. Accordingly, after this initial examination both α and β were used throughout the whole testing phase with their value equal to 0, removing the sparseness and temporal continuity terms from the algorithm. Therefore the cost function used here was the reconstruction term:

$$c(A,Y) = c_T(A,Y). \tag{5.2}$$

5.3. Sound Classification Stage

The classification algorithm used in the last stage of the system was K-NN with $K = 1$, thus 1-NN. This algorithm requires a neighborhood, called a training set, from which it can then compare the distance to a test set. The training set contains the source signals (envelopes) learned by NMF. Therefore, each source signal is a collection of all the points that take part in the envelope. In the same way the test set is also a collection of points. The distance is computed with the square of the Euclidean distance between each point of the test set and the whole assortment of points in the training set.

1-NN will classify and band each point of the test set to a certain class, based on its proximity to certain points of the training set. As mentioned above we have $K = 1$, which means that if the majority of points of a given test sample are nearer the points of a certain training cymbal, then this test sample will be categorized as a sample from that same class [Mitchell 97]. A test sample from a cymbal is considered to be accurately classified if we get above 50% of its source signal points to be assorted as being from that particular cymbal.

To enable an accurate adoption of 1-NN for classification, we first need to transform the test samples intothe same dimensional space as the training set. To do so, the N test samples we feed to 1-NN are first transformed by applying an unmixing matrix A^{-1}, based on the pseudoinverse of the mixing matrix A. This is so that each sample x_i follows the same basis as the training set Y, enabling the test sample to be transformed into y_i, its new representation on the training set's dimensional space:

$$y_i = A^{-1}x_i \;, \quad i \in \{1..N\} \tag{5.3}$$

6. Results and Discussion

This is the chapter we have all been waiting for, the chapter where all the suspense reaches a screeching halt. We finally unveil the results of our study - the good, the bad, and the ugly; the why's and what's.

We start by giving a brief overview of the software and hardware used (section 6.1). We then proceed shedding some light on the ins and outs of the procedures followed while recording cymbal samples in studio, as well as the gear and cymbals used to do so (section 6.2). In the final section of this chapter (section 6.3) we get into the full details on the analysis executed, as well as a full discussion of both the results and decisions taken.

6.1. Hardware and Software Specifications

In this section we will take a look at hardware and software specifications from the tools adopted for this dissertation.

6.1.1. Software Specifications

Analysis Software: Matlab version 7.0.0.19920 (R14).
Operating System: Windows XP with Service Pack 3.
System Type: 32-bit Operating System.

6.1.2. Hardware Specifications

Computer: Asus Notebook F9S Series – bought in 2007.
Processor: Intel Core Duo T7250.

Clock Speed: 2 GHz.

Memory (RAM): 2 GB.

6.2. Cymbal Recording Process

The data used for testing the proposed cymbal classifier was a set of cymbal sounds recorded in the "Chop Chop" studio (Portugal). Figure 6.1 shows a diagram of the studio (the diagram of the studio was provided by the studio owner). Room A is the room where the cymbal samples were recorded, while room B is the control room. Room A has laminated floor, which can easily result in sound wave reflection. To attenuate this effect, the laminated floor is covered with carpets that work as sound absorbers. The walls are made of plasterboard with its interior filled with an acoustic isolator called rockwool. The interior walls of the room, including areas of the ceiling, are covered with sound cushions, which are open structures made of wood and covered with fabric. They house a great quantity of rockwool that work as sound absorbers for the sound waves produced by music instruments. All of this apparatus is of great importance in a recording studio because they prevent reverberation and enable a greater quality in sound control.

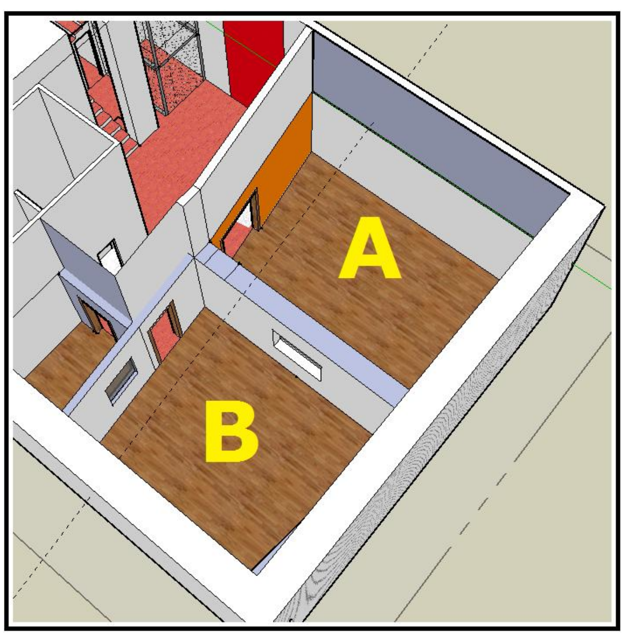


Figure 6.1 – Chop Chop Studio.

The cymbals used for recording were chosen following three criterions - quality, diversity, and sound. Although no attention was given to the cymbals' brands, they are all from top cymbal manufacturers – Zildjian and Sabian Cymbals. It is also important to point out that the number of cymbals available in the studio was limited. We wanted to have at least one quality cymbal for each one of the five classes. We ended up recording six cymbals that are represented on figure 6.2 and listed just below:

- Zildjian 16 inch A Custom Crash Cymbal;
- Zildjian 14 inch K Custom Dark Crash Cymbal;
- Zildjian 16 inch A China High Cymbal;
- Zildjian 9 inch K Custom Hybrid Splash Cymbal;
- Zildjian 14 inch K/Z Hi-Hat;
- Sabian 20 inch AA Heavy Ride.

Figure 6.2 – Cymbals Sampled.
(Top Left) A Custom Crash; (Top Center) K Custom Hybrid Splash; (Top Right) A China High
(Bottom Left) K/Z Hi-Hats; (Bottom Center) AA Heavy Ride; (Bottom Right) K Custom Dark Crash

To play these cymbals we used the signature series drum sticks of the drummer Bruno Pedrosa, made by the Portuguese brand of sticks Missom. These drum sticks are made of "pau-santo". This is not a typical wood for drum sticks manufacturing. They are usually made of maple, hickory, or oak. In figure 6.3 a picture of Pedrosa's stick is shown with the anatomy of a drum stick explained. Other important issues to consider in this whole process regarding the playing techniques utilized while recording the samples. This is a very important point since the sound produced by a cymbal is influenced by the area of the stick it is stricken with. For playing the edges and the bells of the cymbals we used the shoulder of the stick, while for the bow we used the tip. These are the most common areas of the stick for playing those cymbal areas.

Figure 6.3 – Anatomy of a drum stick.

The recorded samples range from the highest level of amplitude to the lowest. The different zones of the cymbal were stricken one by one, from the most powerful of strokes to the softest. Because this process is very susceptive to nuances in the strength used, and the cymbals were hit by hand, this resulted in certain zones of cymbals having more samples than others to ensure we would get a full spectrum of amplitudes. Bell, Bow, and Edge were recorded for four of the six cymbals – A Custom and K Custom Crashes, K Custom Hybrid Splash, and AA Heavy Ride. Due to time restrictions only the edge was used on the first three cymbals and the bow on the ride. As for the china we only recorded and used the edge. For the hi-hat we only got to use the hits on the closed bow.

Table 6.1 shows an overview of the number of samples obtained for each zone of the cymbals considered for this work. In the first column the cymbals family is described, followed by the brand and name of the cymbal. The remaining entries describe the size of the cymbals in inches[23], the zones of the cymbals which were used for analysis, as well as the number of samples per zone and the total number of samples available for each of the

[23] Cymbal sizes are referred in inches amongst drummers, even in Europe.

cymbals' family. All these samples were recorded in mono by a Condenser microphone, an Octava MC012, with a sample rate of 96 kHz.

Cymbal Family	Cymbal	Size	Zone	Number of Samples
Crash	Zildjian K Custom Dark	14 Inches	Edge	23
	Zildjian A Custom	16 Inches	Edge	22
Ride	Sabian AA Heavy	20 Inches	Bow	14
Hi-Hat	Zildjian K/Z	14 Inches	Closed Bow	16
Splash	Zildjian K Custom Hybrid	9 Inches	Edge	20
China	Zildjian Avedis	16 Inches	Edge	20

Table 6.1 – Number of samples available for analyzes.

6.3. Results

For testing, the sampling frequency of each sample was decreased from 96 kHz to 44.1 kHz, due to the size of each sample file, which impaired their use with matlab because of memory constraints. The beginning of each sample was trimmed to assure no silences. The end of each file was also removed, to avoid any unwanted residual sound coming from vibrating metal, which does not contain any distinguishable data about the sound of each class of cymbals.

To obtain the spectrograms of the cymbal samples we used a DFT with 40 millisecond windows and 50% of overlap between them. The length of the DFT was the same as the size of the window. Only the magnitude spectrogram was used, while the phase information was discarded. Several experiments were conducted to analyze our system's ability to separate two and three cymbals with NMF. These experiments are described below.

6.3.1. Two Cymbals

We started by analyzing the system's ability to separate and classify combinations of samples from two cymbals. We used the following combinations:

- Splash Edge with China Edge;
- 14 inch Crash Edge with 16 inch Crash Edge;
- Splash Edge with 16 inch Crash Edge;
- China Edge with 16 inch Crash Edge;
- Hi-Hat Closed Bow with Ride Bow.

These combinations test real situations, especially the combination of both crashes and of the hi-hat with the ride.

To perform this analysis we had to build a training set and a test set. Both types of sets were built with the same combination of two or three cymbals. However, the way we chose the samples for the test set was based on the samples used on the training set:

1. If the training set already has a certain sample, then it will not be used in the test set;
2. In the tests where there are china and hi-hat training sets with low amplitude samples, we used five samples instead of six. This was due to not having enough dynamically spaced (notoriously different amplitude values) samples to work with. By not doing so the training set would become unbalanced, since a certain area of its neighborhood would have more information than the remaining ones;
3. If the training set is comprised of six samples with high amplitudes, then the six test samples will be of low amplitude;
4. If the training set has six low amplitude samples, then the six test set samples will have high amplitude;
5. If the training set has the six samples spread along the various levels of amplitude so does the test sample.

Structuring the samples this way enabled us to analyze our problem from three different perspectives. With (3) we got to simulate situations where the database may only have high amplitude sound files while trying to detect low amplitude samples of cymbal sounds. In (5) we simulated situations where both database and sound sources have a good dynamic range regarding cymbal sounds. The last case (4) was tested just by curiosity.

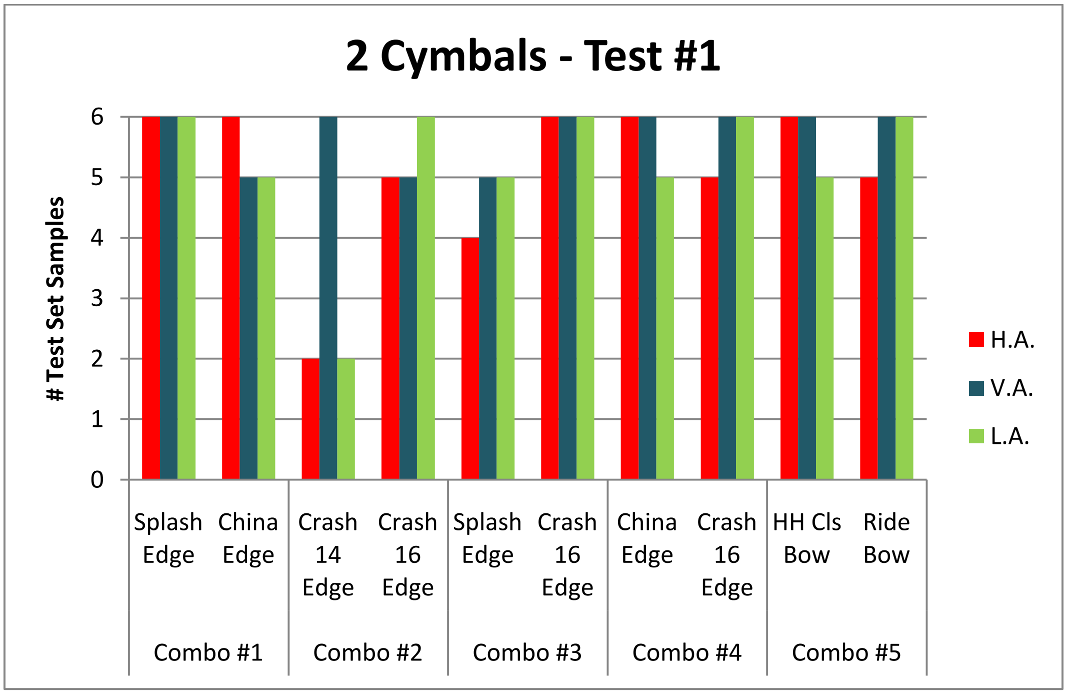


Table 6.2 – Table with the number of correctly classified and separated samples in the first test. H.A. means that the data in the training set consists of high amplitude samples; V.A. is for the training set with variable amplitude samples (i.e., samples that go from high to low amplitude); and L.A. is for a training set with low amplitude samples.

Table 6.2 shows the number of correctly classified samples. The results on this table show that our approach was very successful in identifying the samples correctly. For H.A. we got an overall success of 85% of correctly identified samples, 95% for V.A., and 91,2% for L.A.. A test sample from a cymbal is considered to be accurately classified if in the classification stage we get above 50% of its source signal points to be assorted as being from that particular cymbal. This is done for every test sample of each cymbal used in any combination. The overall success rate values for each of the different types of training sets (H.A., V.A., and L.A) were accounted as the percentage of accurately classified samples between all the combinations with a certain type of training set, against the total number of samples that were tested under that particular type of training set from any cymbal combination. Although the results were very good they were not perfect. So what can possibly be making this happen?

Let us take a look at figure 6.4, which shows the training data (with variable amplitude samples) represented in the space learned by NMF. The training data is from the splash and china cymbals from combination #1 of table 6.2. Most of the points on figure 6.4 that are closer to the *y* axis are from splash samples, while those near the *x* axis are related to the china. Figure 6.5 exhibits an overlap of the source signal points (in green) of a china cymbal test sample (with a variable amplitude training set) over the scatter plot of figure 6.4. This test sample was wrongfully classified on combination #1 (table 6.2), and has the lowest amplitude level amongst the test samples for that particular training set. The points badly classified by 1-NN are exhibited with a blue circle surrounding them. Most of these points are agglomerated in the origin of the coordinate system. Consequently, these points' classification went wrong because of the heavy clustering of training samples' points from both china e splash cymbals near the origin as well. These points get mixed quite easily due to their high mass. So it's only natural that the classification process in this case suffers, ending up by inaccurately classifying the china test sample.

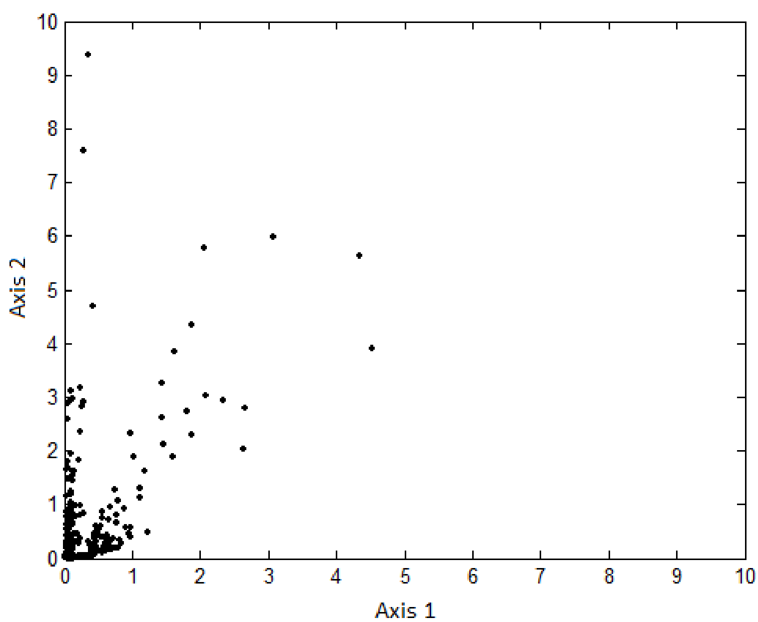


Figure 6.4 – Scatter plot of the training set for V.A. on Combination #1 of table 6.2.

Let us look at one more example, but this time around from combination #3 of table 6.2. For this case we gave use to the last test sample from the splash cymbal of that combination, the sample that was badly classified. In figure 6.6 the points are from the training set of this combination. Taking a look at figure 6.7, we can see the same issue found in figure 6.5. This time however, this problem is responsible for the lack of success in the correct classification of the splash sample on combination #3 of table 6.2 with the lowest amplitude level. All the test samples that were badly classified on H.A. and V.A. had the lowest amplitude levels amongst the samples from the same test set, so there is a pattern here.

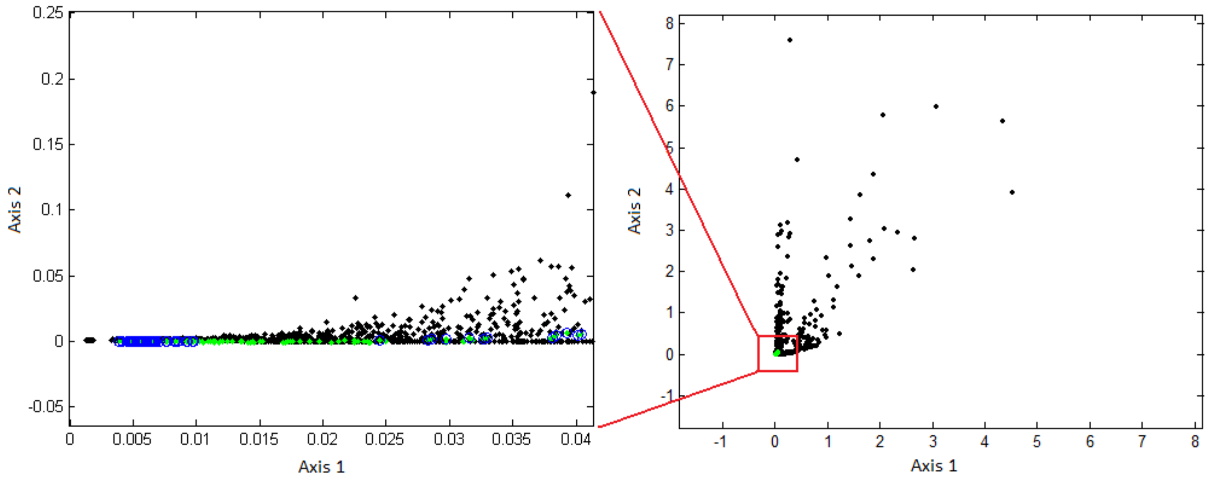


Figure 6.5 – In green the points from the sample with lowest amplitude from the china on combination #1 of table 6.2. The training set has samples with variable amplitudes.

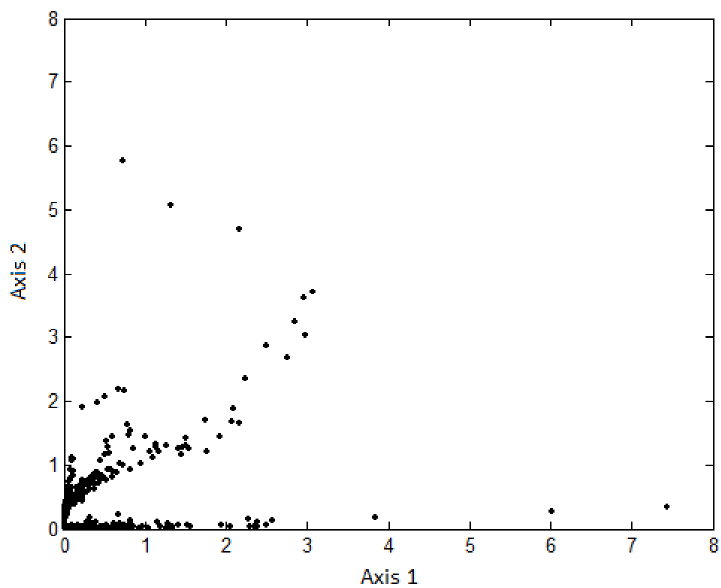


Figure 6.6 – Scatter plot of the training set for V.A. on Combination #3 of table 6.2.

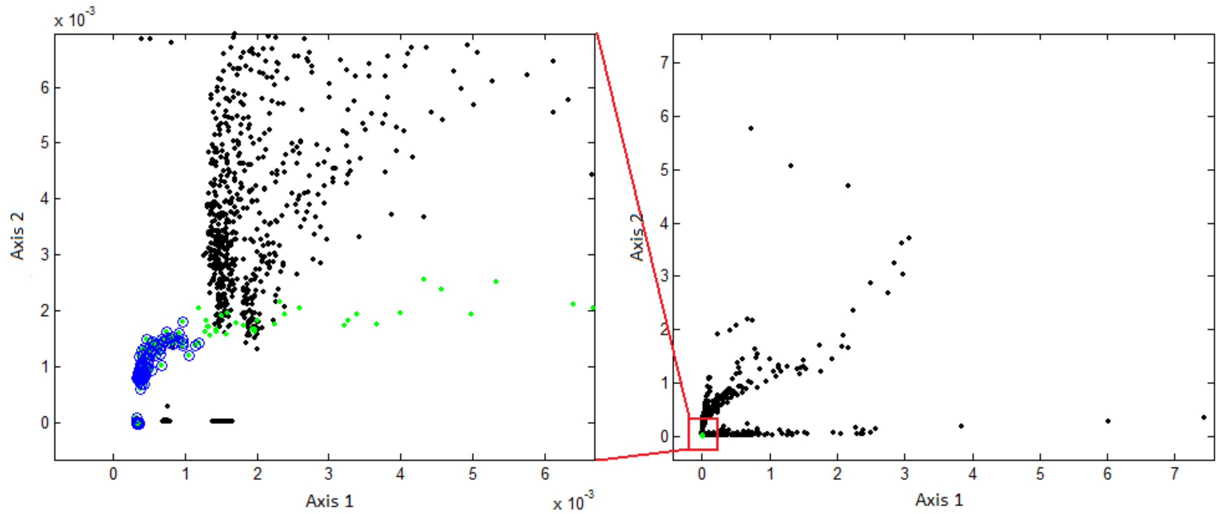


Figure 6.7 – In green the points from the sample with lowest amplitude from the splash on combination #3 of table 6.2. The training set has samples with variable amplitudes.

The training set with low amplitude samples had a different behaviour than the one seen on training sets with variable amplitude and high amplitude samples. None of the badly classified test samples were the lowest in their respective test samples set. However, the classification failure on the test sample from combination #3 originated in the same problem as the one we talked about beforehand. Taking a look at figure 6.8 we can see just that. There is a great mass of points near 0 which end up by being badly classified and inducing an overall wrongful classification of this test sample. Combination #2 has a different problem which we will talk about at the end of this section.

Although the classification in this first test was very good, we took this opportunity to try other approaches to see if we could improve the overall result of the number of samples accurately classified. To do so, we focused our attention on the test sets. From the scatter plots (figure 6.5 and 6.7) we took that the results were not as good as they could be, due to the great quantity of test set points near the origin of the dimensional space, which hailed some inaccurate classifications. To try to overcome this issue we tested two different filters to remove points of a sample with amplitude values below a certain threshold. On figure 6.9 we get to see the filters. Assuming a test sample point is given by $P(x, y)$, then this point is maintained in the test sample only if:

- Right Scatter Plot – **Test #2** - $x > 0{,}06\ OR\ y > 0{,}06$ (if the point is in any of the colored areas);
- Left Scatter Plot – **Test #3** - $x > 0{,}01\ OR\ y > 0{,}01$ (if the point is in any of the colored areas).

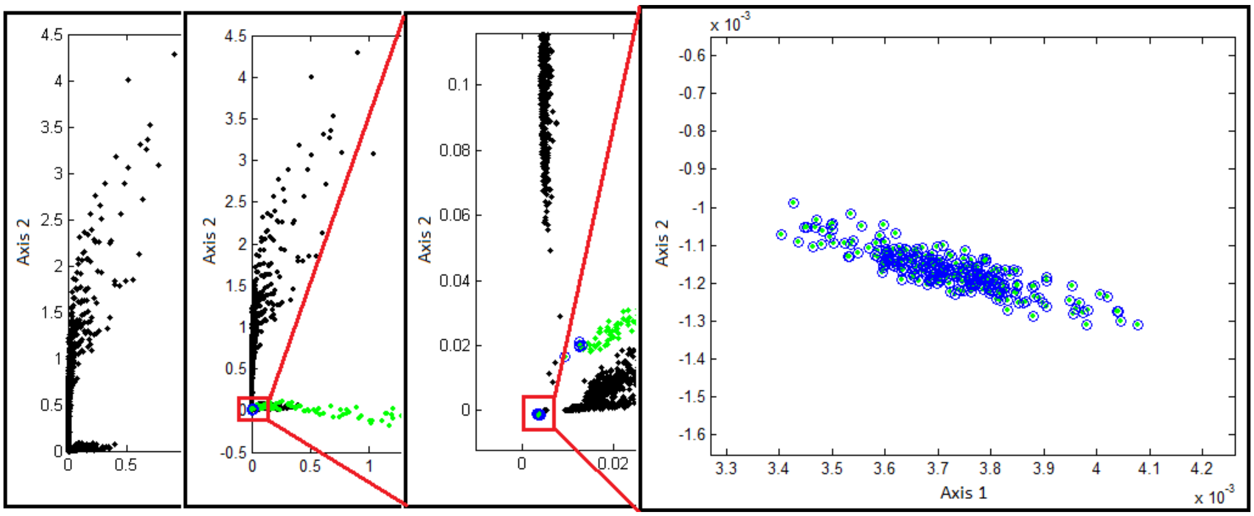


Figure 6.8 – In green, points from the sample of the splash on combination #3 that was badly classified on table 6.2. The training set has samples with low amplitudes.

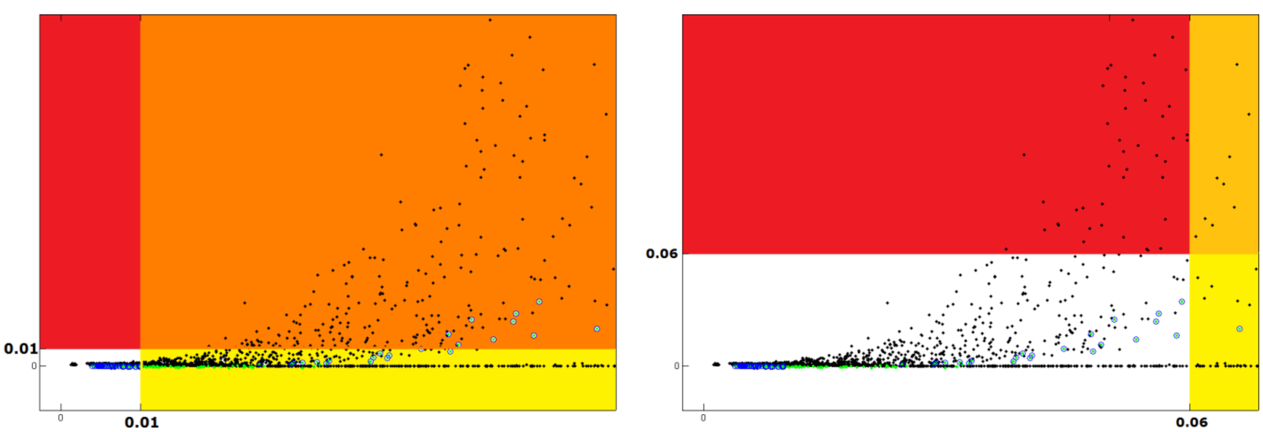


Figure 6.9 – Thresholds.

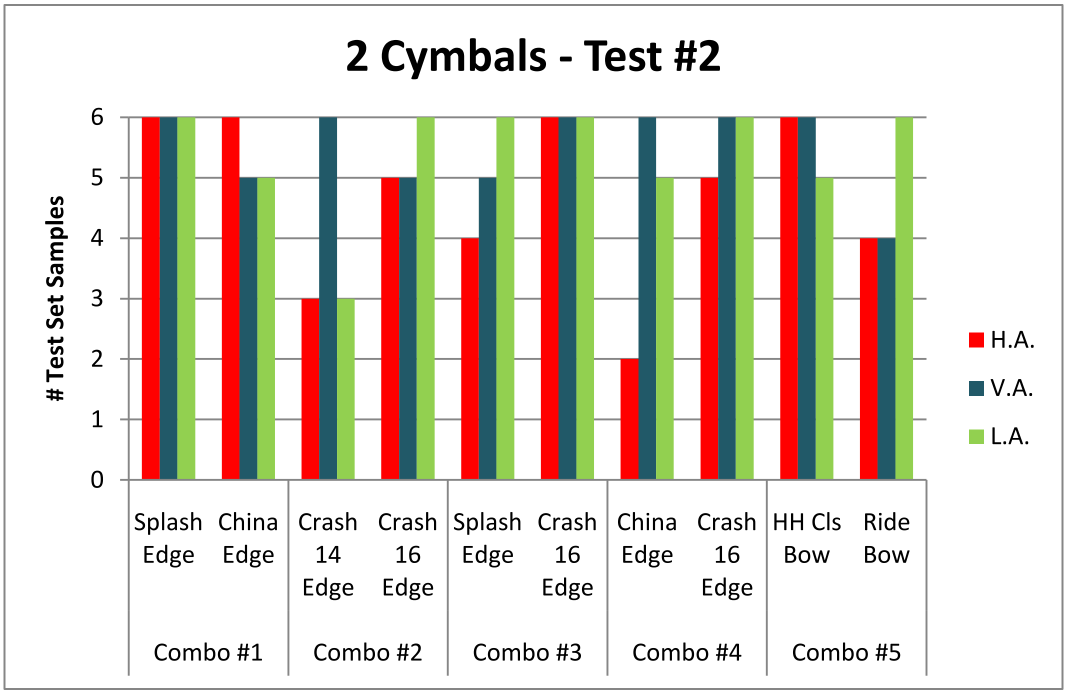


Table 6.3 – Table with the number of correctly classified and separated samples in test #2

These threshold values were chosen at random, since we did not have a way to accurately choose the optimum values for it. Instead of these thresholds we could as well have used another based on the Euclidean distance to the origin. We started by performing test #2, which changed the end result to a certain extent, as can be seen on table 6.3 – it out performs test #1 on combinations #2 and #3, but it is outperformed on combinations #4 and #5. However, this test did not improve the results of the samples that originated this experience at first, and actually aggravated the result of the ride samples of combination #5 with an inferior level of amplitude, and the results from the china on combination #4. It did however improve to 50% the number of 14 inch crash points from a sample, accurately classified on combination #2, when working with H.A. and L.A. training sets.

Seeing test #2 did not produce the results we were expecting, it was decided to try a second approach. This time around instead of using 0,06 as a threshold we used 0,01. The idea behind this change was simple - to get even closer to the origin of the dimensional space. From this change we expected to avoid the loss of unnecessary source signal points, and at the same time improve the overall results. Our suspicious were right, test #3 (table 6.4) did change the end result. Combinations #1, #2, and #5 of test #3 outperform these same combinations on test #1, while combination #4 gets worst. With combinations #1 and #5 this test worked as we were initially expecting it too, with all the samples with the lowest level of amplitude for each test set being accurately classified. However, with combination #3 the number of inaccurately classified samples with the lowest of amplitudes in the test set actually increased.

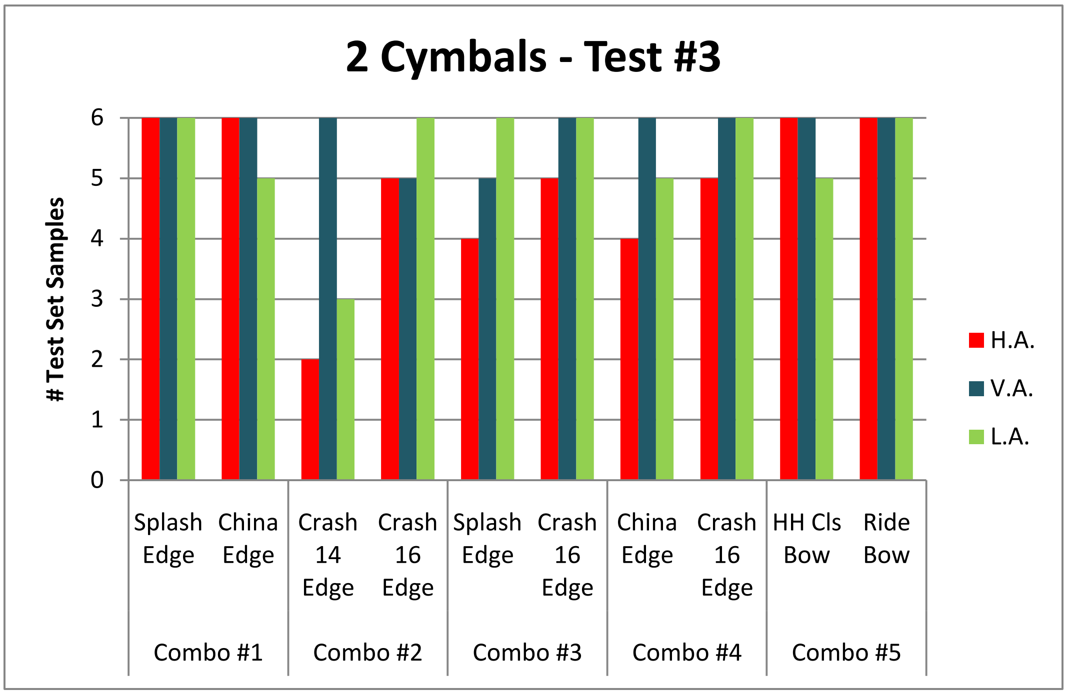


Table 6.4 – Table with the number of correctly classified and separated samples in test #3

The relationship between the basis functions of each cymbal's training samples influences the way the separation is processed. A basis function is not exclusively associated to the sounds of cymbals, being able to find the same properties seen on different cymbals. This prevents the separation from being perfect since we will be having information from each cymbal on each basis function. If this was not the case, the results of classification would probably be of 100%, since we would have a basis for each cymbal. However this was not the case hence the results we got. Thus the quality of the classification depends on NMF's ability to accurately separate the sound sources and the basis for each cymbal from the mixture.

To have a better idea of how NMF really affects the results of classifying samples with 1-NN, next, we take a look at the source signals separated from the mixtures shown on table 6.2, of combinations #1 – training set with H.A. samples (figure 6.10), combination #2 – training set with H.A. samples (figure 6.11) and combination #4 – training set with V.A. samples (figure 6.12). Each figure shows the source signals learned by NMF, where a source signal is a temporal envelope that contains the coefficients related to one spectrogram (from the training set) and one basis function. The squares mark the peaks of the envelopes associated to the samples for each cymbal – one color for each cymbal.

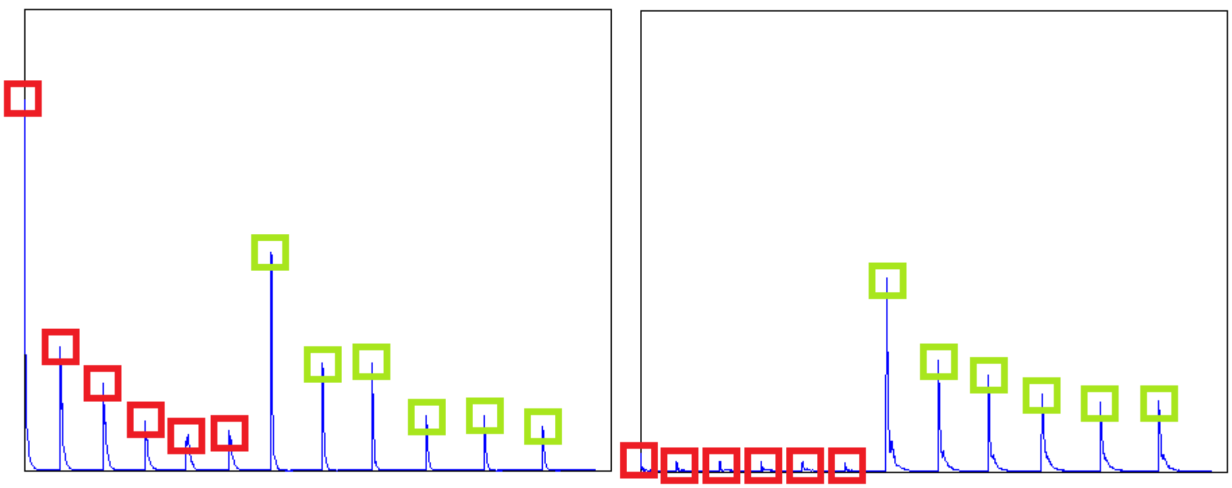


Figure 6.10 – Source signals from splash (left) and china (right) obtained by NMF, with a training set with high amplitude samples. The first 6 source signals (envelopes) in each figure are related to the splash cymbal while the other 6 are related to the china. The squares mark the peaks of the envelopes associated to the samples for each cymbal – red for splash and green for china. Any figure shows the source signals related to one of the basis functions learned by NMF.

On figure 6.10 it is shown that the left diagram has very strong elements from both cymbals. Since in the right figure the peaks from splash samples are much lower than those from china samples, it seems that the basis function related to these source signals is describing properties from the china cymbal. On figure 6.11 we have the same thing happening, but instead of a splash and a china we have a 14 inch crash and a 16 inch crash. The way NMF learned these source signals seen on both figures (6.10 and 6.11) is the result of the different combinations of types of training sets we made with the different types of test sets. From the three types of training sets of table 6.2, the ones with samples that range from a high level of amplitude to a low level are the ones that show the best results in all the combinations. This is shown on table 6.5. The table with variable amplitude training samples has the highest percentage of accurately classified samples. This is so because the test sets have variable amplitude sample also. This way, NMF will have enough information in the training set to accurately recognize each of the test samples. This is especially shown on combination #2 of table 6.2. The cymbals on that combination are of the same class - crash cymbal. Thus, it is expected of them to have very similar characteristics, which is assumed to bring about problems when NMF tries to separate them into two different source signals. However, due to the variable amplitude training and test sets we get eleven out of twelve accurately classified samples. The tests with high amplitude and low amplitude training samples do not have enough information in them to give NMF the tools to better separate the low amplitude and high amplitude tests sets respectively.

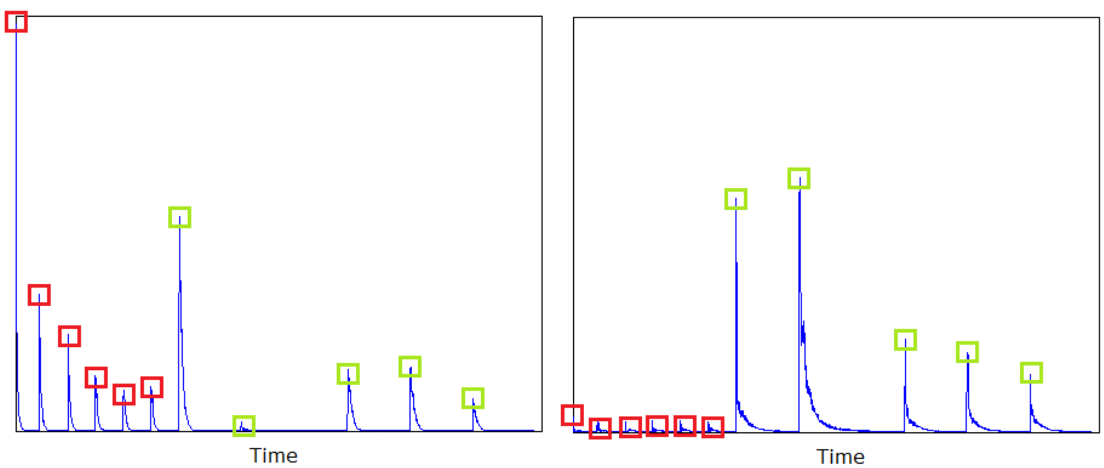


Figure 6.11 – Source signals from 14 inch crash (left) and 16 inch crash (right) obtained by NMF, with a training set with high amplitude samples. The first 6 source signals (envelopes) in each figure are related to the 14 inch crash cymbal, while the other 6 are related to the 16 inch crash. The squares mark the peaks of the envelopes associated to the samples for each cymbal – red for 14 inch crash and green for the 16 inch crash. Any figure shows the source signals related to one of the basis functions learned by NMF.

		S.R. (%)	C.S. (#)
Combo #1	Splash Edge	91,3	6
	China Edge	80,8	6
Combo #2	Crash 14 Edge	54,3	2
	Crash 16 Edge	67,9	5
Combo #3	Splash Edge	64,4	4
	Crash 16 Edge	91,7	6
Combo #4	China Edge	75,9	6
	Crash 16 Edge	84,9	5
Combo #5	HH Cls Bow	100	6
	Ride Bow	81,2	5
	Average (%)	79,24	85
	Total		51/60

		S.R. (%)	C.S. (#)
Combo #1	Splash Edge	98,8	6
	China Edge	61,4	5
Combo #2	Crash 14 Edge	79,2	6
	Crash 16 Edge	75,1	5
Combo #3	Splash Edge	73,5	5
	Crash 16 Edge	99,6	6
Combo #4	China Edge	89,3	6
	Crash 16 Edge	99,9	6
Combo #5	HH Cls Bow	100	6
	Ride Bow	87,1	6
	Average (%)	86,39	95
	Total		57/60

		S.R. (%)	C.S. (#)
Combo #1	Splash Edge	93,2	6
	China Edge	68,6	5
Combo #2	Crash 14 Edge	49,7	2
	Crash 16 Edge	96,6	6
Combo #3	Splash Edge	87,2	5
	Crash 16 Edge	100	6
Combo #4	China Edge	93,2	5
	Crash 16 Edge	99,9	6
Combo #5	HH Cls Bow	99,8	5
	Ride Bow	99,1	6
	Average (%)	88,73	91,2
	Total		52/57

Table 6.5 – Combinations with high amplitude training sets (left table). Combinations with variable amplitude training sets (center table). Combinations with low amplitude training sets (right table).The success rate (S.R.) is the percentage of source signal points to be assorted as being from a particular cymbal. The column with the number of correct samples (C.S.) shows the number of accurately classified samples for each cymbal in any combination. Average gives the average of the S.R. over all the cymbals in each combination, and it also gives the percentage of accurately classified samples over all the cymbals in the combinations. Total represents the total number of C.S. samples over the total number of samples testes over all the cymbals.

We have shown here how the proposed classifier achieves very accurate results when it comes to cymbal classification. In addition, we have analyzed the badly classified cases. We were able to conclude that the quality of the classification depends on NMF's ability to accurately separate the sound sources from mixtures, and that the training sets with samples that range from a high level of amplitude to a low level are the ones that show the best results in all the combinations, that is, if the test sets also have well distributed samples amongst amplitude levels. We also experimented with the usage of test sample source signal points' filters that ended up improving the end result of our tests to some extent. Since the results were very good on the classification of two cymbals, we took our analysis a step further to see how the classifier behaved when the signal mixture was composed of samples from three cymbals. This analysis is discussed in the next section.

6.3.2. Three Cymbals

The next step in our analysis was of performing the same three tests performed with two cymbals, but this time around with a combination of three cymbals. We did not perform the same amount of combinations as we did on two cymbals due to time restrictions. The combinations chosen are as follows:

- Splash Edge, 16 inch Crash Edge, and China Edge;
- Splash Edge, 16 inch Crash Edge, and 14 inch Crash Edge;
- China Edge, 16 inch Crash Edge, and 14 inch Crash Edge;

Like with what we saw on section 6.3.1. and for the same reasons, the china cymbal in the tests with low amplitude training sets has only five samples in the training and test sets. The rules for building the test and training sets are the same as what we saw on the previous section.

Table 6.6 shows the number of correctly classified samples for all the combinations of three cymbals we analyzed. In it, it is shown that our approach was successful in identifying the samples correctly. For H.A. we got an overall success of 66,6% of correctly identified samples, 74,1% for V.A., and 90,3% for L.A. A test sample from a cymbal is considered to be from a cymbal *X* if in the classification stage it has a greater percentage of source signal points from cymbal *X*. This is done for every test sample of each cymbal used in any combination. The overall success rate values for each of the different types of training sets H.A., V.A., and L.A were accounted as the percentage of accurately classified samples between all the combinations with a certain type of training set, against the total number of samples that were tested under that particular training set from any cymbal combination.

With V.A. and L.A. training samples we get above 50% of success with combinations #2 and #3, and #1 and #2 respectively. The main issue here is with the china on combo #1, which shows very low results with V.A. and H.A. training samples. Just like with two cymbals we tested two different filters to remove points from a test sample with amplitude values below a certain threshold. Figure 6.9 from the previous section shows how the two filters work.

Tables 6.7 and 6.8 show the results of applying said filters to the same combos we saw on test #1 of this section.

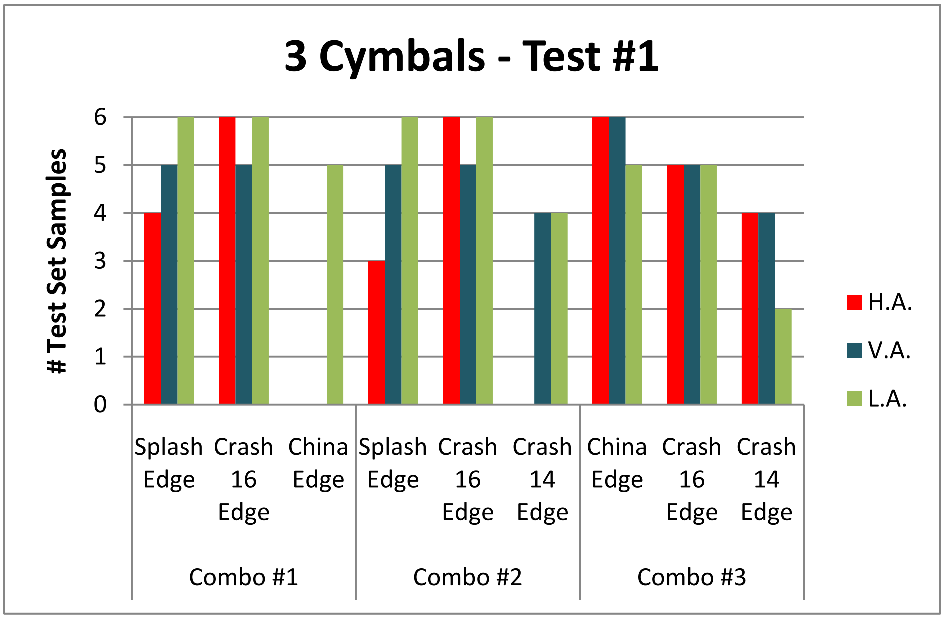


Table 6.6 – Table with the number of correctly classified and separated samples in the first test. H.A. means that the data in the training set consists of high amplitude samples; V.A. is for the training set with variable amplitude samples (i.e., samples that go from high to low amplitude); and L.A. is for a training set with low amplitude samples.

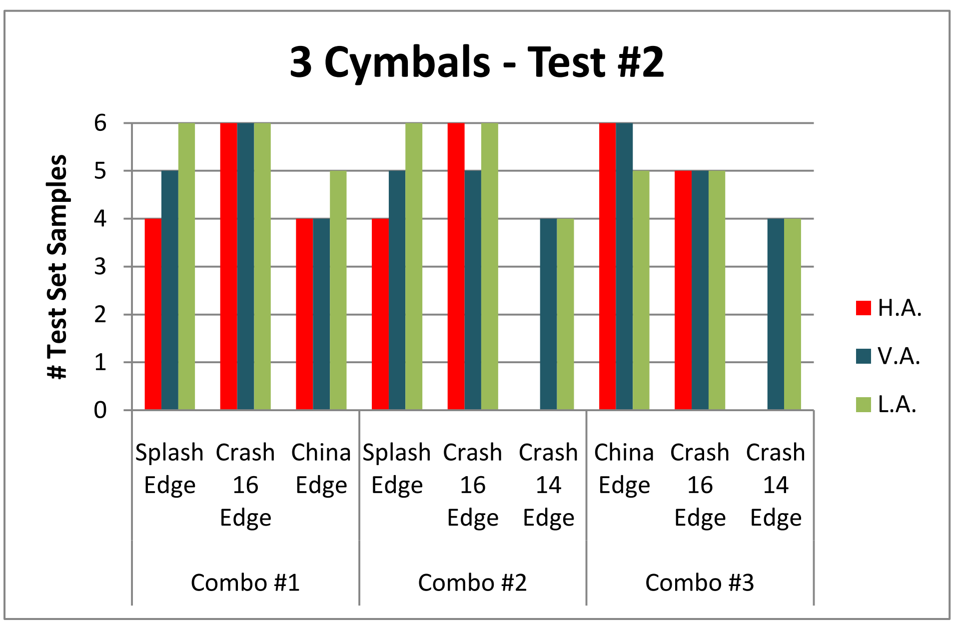


Table 6.7 – Table with the number of correctly classified and separated samples in the second test, with threshold <= 0,06.

From the two extra tests, only the test with threshold <= 0,06 (table 6.7) shows really great improvements when compared to test #1 (table 6.6). For H.A. we got the same amount of successfully indentified points (66,6%), while with V.A. and L.A. the results improved to 83,3% and of 94,2% respectively. All tests with V.A. and L.A. training sets had success above 50%, which was a great improvement from test #1. Even the china on combination #1 improved greatly with above 50% of success with H.A., V.A., and L.A. training sets.

Testing the classification of three cymbals was done in the exact same conditions as what we saw with two cymbals. However, while with two cymbals using the extra tests (#2 and #3) hailed some improvements, in this case we got a very good improvement with test #2 and a lighter improvement with test #3. Another surprising result came from the two combinations of three cymbals that contain the 14 inch crash and the 16 inch crash. In whatever test the results were very good with these two cymbals included. There was one other surprising result here. While with two cymbals the variable amplitude training and test sets hailed the best results, here the best results came from the low amplitude training sets with high amplitude test sets.

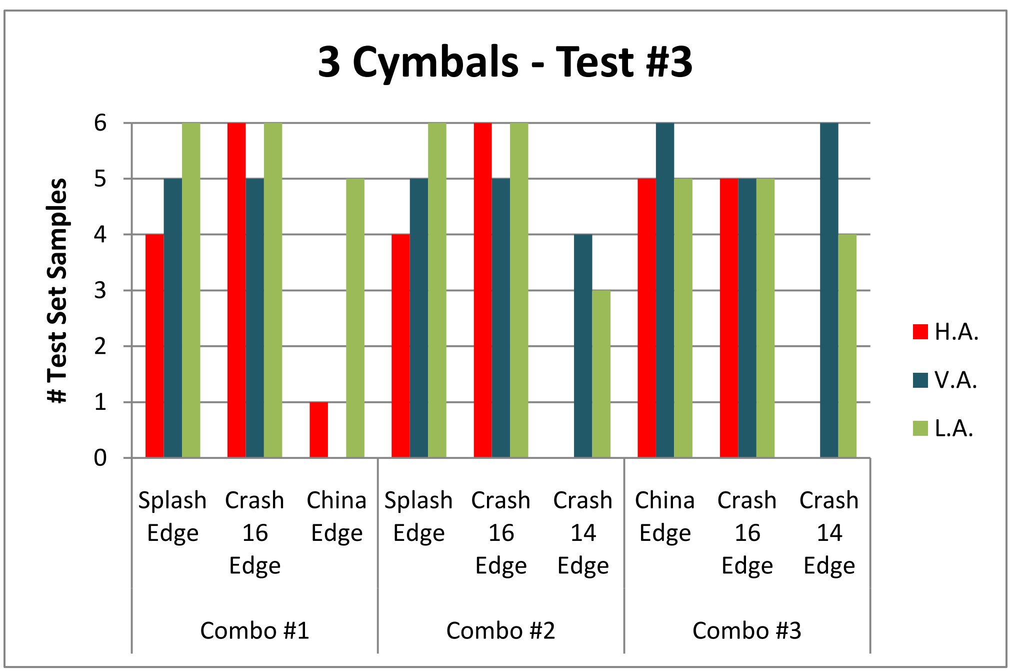


Table 6.8 – Table with the number of correctly classified and separated samples in the third test, with threshold <= 0,01.

The results in this section are once again adamant in showing that the combination of NMF and 1-NN enable a great level of success when it comes to separating samples from cymbals. Taking in consideration we are handling three cymbals instead of two, makes these results that more regal. This experiment also showed how the usage of filtering can be of great importance to improve the accurate classification of cymbals.

For checking the results of the separation and classification with farther detail, check the Attachments. There the reader can find two sections, each one with the entire collection of tables whose values were taken from the different experiments made with two and three cymbals.

7. Conclusions

The idea of transcribing a piece of music in its most detailed shape, with rhythmic, harmonic, and melodic content, is as incredible as useful. From learning music on an instrument, creating music applications, to software able to enhance the work of DJs; the possibilities are immense. However, before even thinking about transcription we have to first contemplate how to accurately classify the instruments in a musical piece. Correct classification is the first step for achieving a precise transcription. Most proposed classifiers of musical instruments deal with string and wind harmonic instruments, while much less attention has been given to percussion instruments with non-perceptible pitch, that is, with indefinite pitch. The classification of cymbal events, an area which as far as we know as never been tackled in the scientific world before, presents itself as challenging. This is due to the very noisy spectrum these percussive instruments have. To separate cymbals from one another is a very complex task, since there is not a clean a definite spectrum like the ones on pianos and flutes for instance. The goal of this dissertation was to explore automatic cymbal classification and the identification of which class of cymbals (crash, ride, splash, china, and hi-hat), cymbals played belong to.

We were able to achieve a great level of success by accurately classifying various combinations of two or three cymbals played sequentially. To achieve this goal we had to create a training set of samples for each cymbal in the sequence. This set would then have a sound classifier be applied to it. The choice of an adequate technique is one of the first problems one encounters. Whereas most sound classifiers use a set of pre-defined features [Bilmes 93][Gouyon 01][Herrera 02][Kaminskyj 01][Tindale 04][Schloss 85][Sillanp 02], there are also some classifiers that learn the features using a decomposition method [Abdallah 03][FitzGerald 04][Hélen 05][Paulus 05]. In this work we reviewed several of these decomposition techniques and worked with three of them – PCA, ICA, and NMF. As we had predicted, PCA due to its constraints did not give satisfactory results. ICA's results were also not very satisfactory, so we decided to focus our attention on NMF. It could have very well been PSA for that matter, since it also seemed to guaranty good results. But NMF has something very special about it. It represents data in a parts based approach; it deconstructs information into non-negative parts which when summed up give the whole once again. This is a very natural way of approaching classification since that is what we humans do, we can deconstruct the sound mixture into the various instruments – guitar, cymbals, piano, snare, bass drum, while we listen to it. After deconstructing the original signal mixtures from the training set into various source signals, we can proceed to classify new data samples. The source signals are the values of the basis functions, which are the features. For classifying the data samples we chose to use 1-NN. This algorithm classifies new data samples based on their proximity to the points in the training set.

Here we proved that a combination of NMF with 1-NN is a good option for automatic cymbal classification. For testing this model we assembled five different combinations of two cymbals and three different combinations of three cymbals played consecutively. For each combination we had three different collections of 5 or 6 samples as a training set, and 6 other samples as a test set. Our classifier achieved excellent results for sound mixtures with these combinations. The quality of the classification was proportional to the quality of the separation, i.e., the higher the quality of the sound source separation done by NMF, the higher the success of classification. The overall classification rate for each of the three collections of training samples for all the combinations of two cymbals was of 85%, 95%, and 91,2%. The most surprising results came from the combination of the ride bow and closed hi-hat bow, given these cymbals have similar characteristics – both have very low energy and a fast decay. The overall classification rate for each of the three collections of training samples for all the combinations of three cymbals was of 66,6%, 74,1%, and 90,3%. The most surprising results in this case, came from the combinations which contained both the 14 inch and the 16 inch crash. Given these cymbals are of the same family and have similar characteristics, the result were very good. We were able to prove that a combination of NMF with 1-NN is a good option for automatic cymbal classification.

Although the classifications were very good, we took this opportunity to try other approaches to see if we could improve the overall result of the number of samples accurately classified. For the tests with two cymbals, the approaches did not improve the results that much, although the test with a threshold of 0,01 showed the best overall results. For three cymbals one of the approach with a threshold of 0,06 had a definite positive impact on the overall success rate of the classification of cymbal samples, with results of 66,6%, 83,3%, and 94,2%, which improved upon the initial results.

7.1. Future Work

PSA is also a good candidate for performing an accurate sound source separation just like NMF. So the next natural step to follow in this study, would be that of using PSA instead of NMF for sound source separation.

During the testing phase of this work we also dabbled with the usage of the Mahalanobis distance in K-NN, instead of using the Euclidean distance. We did so because the Mahalanobis distance is a technique for calculating the distance between two points that is better adapted than the Euclidian distance to settings involving non spherically symmetric distributions, which is the case of our subject of study. However, we were not able to go really deep into its possibilities due to time constraints. So a further study of the Mahalanobis distance with K-NN could be very promising.

Also, even though we used 1-NN in the final stage of our classifier, many other algorithms are worth considering, like k-means or support vector machines (SVM). It would also be interesting to see how this setup would work with other zones from cymbals that we didn’t work with, like crash and ride bell, open hi-hat, hi-hat foot chick and foot-splash.

To shed some more insight into how the cymbals may affect the outcome of the sound source separation stage with NMF, we decided to study some frequency and envelope characteristic of cymbals. This would be a complement to a study we did on cymbals’ physical characteristics; like the way size, material, and shape (just to name a few) are relevant in modeling the sound and various frequencies of a cymbal when stricken. Since this was not the main focus of our work, we were forced to drop this analysis due to time constraints. Nonetheless, this is an important study to understand the main sound characteristics that really drive the timber and frequencies of each class of cymbals, which in turn can help in understanding how samples can be manipulated to improve the performance of classification. We feel this study about the instruments would also be very beneficial in developing a general procedure for anyone who may want to record samples of their own, or even for developing a complete and general scientific samples database. This general procedure would also be important in setting rules for the types of stick to use. The type of stick, size, shape of tip, weight of the stick, type of wood, etc. all influence the final sound that comes out of a cymbal.

8. References

[Abdallah 03] Abdallah, S.A. and Plumbley M.D., "An Independent Component Analysis Approach to Automatic Music Transcription" in Proceedings of the 114th Audio Engineering Society Convention, Amsterdam, Netherlands, March 22-25, 2003.

[Arons 92] Arons, B., "A Review of the Cocktail Party Effect" in Journal of the American Voice I/O Society, Vol. 12, pp. 35-50, July 1992.

[Attneave 54] Attneave, F, "Some Informational Aspects of Visual Perception" in Psychological Review, Vol. 61, No. 3, pp.183-193, May 1954.

[Barlow 59] Barlow, H.B., "Sensory mechanisms, the reduction of redundancy, and intelligence" in NPL Symposium on the Mechanization of Thought Process, No. 10, pp. 535-539, HM Stationery Office, London, 1959.

[Barlow 01] Barlow, H.B., "Redundancy Reduction Revisited" in Network: Computation in Neural Systems, Vol. 12, No. 3, pp. 241-253, March 2001.

[Bilmes 93] Bilmes, J., "Timing is of the essence: Perceptual and computational techniques for representing, learning and reproducing expressive timing in percussive rhythm," M.S. thesis, Massachussetts Institute of Technology, Media Laboratory, 1993.

[Berg 95] Berg, R.E. and Stork D.G., "The Physics of Sound – Second Edition", Prentice Hall, 1995.

[Burrows 99] Burrows, T., "How to Read Music: Reading Music Made Simple", St. Martin's Press, New York, 1999.

[Burrus 08] Burrus, C.S., Frigo, M., Johnson, S.G., Pueschel, M. and Selesnick I., "Fast Fourier Transforms" in Connexions Website, Rice University, Houston, Texas, 2008.

[Byrd 02] Byrd, D. and Crawford, T., "Problems of Music Information Retrieval in the Real World" in Information Processing and Management: an International Journal, Vol. 38, Issue 2, pp. 249 – 272, March 2002.

[Cavaco 07] Cavaco, S. and Lewicki, M.S., "Statistical Modeling of Intrinsic Structures in Impacts Sounds" in Journal of the Acoustical Society of America, Vol. 121, Issue 6, June 2007.

[Cavaco 09] Cavaco, S., "Lecture 1 – Introduction and Audio Analysis" in Interaction and Visualization in Multimedia Environments – Audio, FCT-UNL (New Faculty of Science and Technology of the New University of Lisbon), March 26, 2009.

[Chen 07] Chen, C.H., "Lecture 13 - Cost Functions" in Principles of Microeconomics, Massachusetts Institute of Technology, October 15, 2007.

[Cohen 95] Cohen, L., "Time-Frequency Analysis", Englewood Cliffs, NJ Prentice Hall, 1995.

[ECE 10] ECE 410: Digital Signal Processing – Short Time Fourier Transform; Video's content adapted from ECE 410: Digital Signal Processing course notes by David Munson and Andrew Singer, webpage: http://www.youtube.com/watch?v=TZzS52OplYs, accessed on 5 of June of 2010;

[Eco 98] Eco, U., (1977), "Como se Faz uma Tese em Ciências Humanas", Ed. Presença, 1998.

[Everest 01] Everest, F.A., “The Master Handbook of Acoustics - Fourth Edition”, McGraw-Hill, 2001.

[FitzGerald 02] FitzGerald, D., Coyle, E. and Lawlor, B., “Sub-Band Independent Subspace Analysis from Drum Transcription” in Proceedings of the 5th International Conference on Digital Audio Effects (DAFX-02), Hamburg, Germany, September 26-28, 2002.

[FitzGerald 03a] FitzGerald, D., Lawlor, B. and Coyle, E., “Prior Subspace Analysis for Drum Transcription”, in Proceedings of the 114th Audio Engineering Society Convention, Amsterdam, Netherlands, March 22-25, 2003.

[FitzGerald 03b] FitzGerald, D. Coyle, E. and Lawlor, B., “Independent Subspace Analysis Using Locally Linear Embedding” in Proceedings of the 6th International. Conference on Digital Audio Effects (DAFX-03), London, UK, September 08-11, 2003.

[FitzGerald 04] FitzGerald, D., “Automatic Drum Transcription and Source Separation” Doctoral Thesis, Dublin Institute of Technology, 2004.

[Gerou 96] Gerou, T. and Lusk, L., “Essential Dictionary of Music Notation”, Alfred Publishing Co., 1996.

[Ghias 95] Ghias, A., Logan, J., Chamberlin, D. and Smith, B.C., “Query By Humming: Music Information Retrieval in An Audio Database,” in Proceedings of the third ACM international conference on Multimedia, pp. 231-236, San Francisco, California, United States, November 05-09, 1995.

[Gouyon 01] Gouyon , F., and Herrera, P., “Exploration of techniques for automatic labeling of audio drum tracks’ instruments,” in Proceedings of MOSART: Workshop on Current Directions in Computer Music, 2001.

[Hélen 05] Hélen, M. and Virtanen T., “Separation of Drums from Polyphonic Music Using Non-Negative Matrix Factorization and Support Vector Machine”, in Proceedings European Signal Processing Conference, Istanbul, Turkey, 2005.

[Herrera 02] Herrera, P., Yetarian, A., and Gouyon F., “Automatic classification of drum sounds: A comparison of feature selection and classification techniques,” in Proceedings of the International Conference on Music and Artificial Intelligence, 2002, pp. 79–91.

[Ientilucci 03] Ientilucci, E. J., “Using the Singular Value Decomposition”, Chester F. Carlson Center for Imaging Science, 2003.

[Kapur 04] Kapur, A., Benning, M., and G. Tzanetakis, “Query by Beatboxing: Music Information Retrieval for the DJ” in Proceedings of the International Conference on Music Information Retrieval, Barcelona, Spain, October 11-14, 2004.

[Kaminskyj 01] Kaminskyj, I., “Multi-feature musical instrument sound classifier,” in Proceedings of Australasian Computer Music Conference, 2001.

[Lee 99] Lee, D.D. and Seung, H.S., “Learning the Parts of Objects by Non-Negative Matrix Factorization” in Nature 401, pp.788–791, October 21, 1999.

[Lee 01] Lee, D. D. and Seung, H. S., “Algorithms for Non-Negative Matrix Factorization” in Neural Information Processing Systems, pp. 556-562, Denver, 2001.

[Miranda 07] Miranda, A. A., Borgne, Yann-Aël Le and Bontempi, G., “New Routes from Minimal Approximation Error to Principal Components” in Neural Processing Letters, Springer, Vol. 27, No. 3, pp. 197-207, June, 2008.

[Mitchell 97] Mitchell Tom M., “Machine Learning”, McGraw-Hill, March, 1997.

[Moreau 07] Moreau, A. and Flexer, A., “Drum Transcription in Polyphonic Music Using Non-Negative Matrix Factorisation” in Proceedings of ISMIR, 2007.

[Olshausen 96] Olshausen B.A. and Field D.J., “Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images” in Nature 381, pp.607-609, June 13, 1996.

[Paulus 05] Paulus, J and Virtanen T., “Drum Transcription With Non-Negative Spectrogram Factorisation” in Proceedings of European Signal Processing Conference, Turkey, 2005.

[Paulus 06] Paulus, J., “Acoustic Modelling of Drum Sounds with Hidden Markov Models for Music Transcription” in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vol.5, Issue 14-19 May, pp. 241-244, July 24, 2006.

[Pinksterboer 92] Pinksterboer, H., “The Cymbal Book”, Hal Leonard Corporation, 1992.

[Plumbley 02] Plumbley, M.D., Abdallah, S.A., Bello, J.P., Davies M.E., Monti G. and Sandler M.B., “Automatic Music Transcription and Audio Source Separation” in Cybernetics and Systems, Vol. 33, Issue 6, pp 603-627, September, 2002.

[Sam 08] Sam, M. and Pritts R., “Audio Sampling: A Practical Guide”, Focal Press, 2008.

[Schloss 85] Schloss, W., “On the automatic transcription of percussive music -from acoustic signal to high-level analysis”, Ph.D. thesis, CCRMA, Department of Music, Stanford University, 1985.

[Shea 05] Shea, M., “Studio Recording Procedures – How to Record Any Instrument”, McGraw-Hill, 2005.

[Sillanp 02] Sillanp, J., “Drum stroke recognition,” Tech. Rep., Tampere University of Technology, 2002.

[Simon 08] Simon, I., Morris, D. and Basu, S., “MySong: Automatic Accompaniment Generation for Vocal Melodies” in Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, pp. 725-734, Florence, Italy, 2008.

[Smaragdis 03] Smaragdis, P. and Brown, J.C., “Non-Negative Matrix Factorization for Polyphonic Music Transcription” in Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 177-180, October 19-22, 2003.

[Smith 02] Smith, L.I., “A Tutorial on Principal Component Analysis”, Cornell University, United States of America, February 26, 2002.

[Stone 04] Stone, J., “Independent Component Analysis, A Tutorial Introduction”, MIT Press, 2004.

[Ryynänenm 08] Ryynänenm, M., Virtanen, T., Paulus, J. and Klapuri A., "Accompaniment Separation and Karaoke Application Based on Automatic Melody Transcription" in Proceedings of IEEE International Conference on Multimedia & Expo (ICME'08), pp. 1417-1420, Hannover, Germany, June 23, 2008.

[Tindale 04] Tindale, A., Kapur, A. and Fujinaga I., “Towards Timbre Recognition of Percussive Sounds”, 2004.

[Virtanen 03] Virtanen T., “Sound Source Separation Using Sparse Coding With Temporal Continuity Objective” in Proceedings of International Computer Music Conference (ICMC2003), Singapore, 2003.

[Virtanen 07] Virtanen, T., “Monaural Sound Source Separation by Nonnegative Matrix Factorization”, in IEEE Transactions on Audio, Speech and Language Processing, Vol.15, No. 3, pp.1066 – 1074, March, 2007.

[Widrow 61] Widrow, B., “Statistical Analysis of Amplitude-Quantized Sampled-Data Systems” in AIEE Transactions on Applications and Industry, pp. 1 – 14, January, 1961.

. Attachment #1

. A Bit of History

All the information in this section is taken from Pinksterboer's book [92] about cymbals.

Cymbals are not like any other instrument, in that they are used in almost every style of music - jazz, marching band, orquestral, rock, Afro-Cuban, heavy-metal; the list could really go on. However, it is believed that cymbals may have come from a very different background regarding its usage.

Bronze is the oldest alloy known to man, and the natural resource that has always been adopted for cymbal making. In has been used in Asia since around 3000 B.C. (before Christ), so cymbals' ancestors may have been from that time. Nonetheless, one of the first stories known about cymbals dates back to 1200 B.C. where the worshiping of the goddess Cybele was always accompanied by the sound of cymbals. In the holy bible the first reference to a cymbal dates back to 1050 B.C., when David moved the Ark of God to Jerusalem, and at his arrival[24]:

> *… and all the house of Israel played before the Lord on all manner of instruments made of Firrewood, even on harpes and on psalteries, and on timbrels, and on cornets, and on cimbels.*

Still, the usage of these metallic saucers was not exclusively reserved to worshiping Gods; they were used in a numerous ceremonies and parties, including orgies and funerals, while witches used them to counter lunar eclipses.

[24] The following text is a transcription from what is stated on the Bible. The way it was written, although wrong from today's standards, has to be respected. Thus, it isn't filled with typos.

The military also found usages for cymbals. They were an integral part of the military music of the Turkish army during the Ottoman Empire. The Chinese army of about 2500 years ago used them to strike terror in their enemy lines with a cacophony of clashing cymbals, a technique that appears to have been used also in the Korean war of the 1950's. The European military marching bands have also been using cymbals since the eighteenth century.

It was not until the second half of the nineteenth century that cymbals started to be used widely as a serious musical instrument, mainly due to the extensive cymbal parts of authors like Wagner and Berlioz. The latter was also the first to require the cymbals to be suspended and played with wooden sticks. This was a big thing at the time, because cymbals were used in pairs, with each one attached to a any hand of the percussionist. The musician would then clash the cymbals against each other. Verdi and Rossini were fundamental in continuing the development of the usage of the cymbal and of creating the most used technique in a contemporary setting, when it comes to a drum kit player - combining the stroke of a cymbal with a bass drum hit. With the advent of the bass drum pedal and the inception of the drum kit cymbals started to garner more attention from musicians, and it was at the beginning of the last century that the trends that are followed nowadays started to be developed.

. Drum Kit Sound Recording and Production

The idea behind this section is that of giving a very brief insight into drum set recording methods. This way, it becomes easier to understand some of the options taken when trying to perform sound source separation, like using spectrograms with algorithms like ICA.

When music is recorded in a studio a great number of microphones is usually required. Each instrument can have more than one microphone assigned to the recording of its sound. A drum kit is a collection of percussion instruments, making it a very special instrument regarding music production. The techniques utilized for the recording of a kit, as well as the placement and number of microphones, vary accordingly to a great number of factors, like the number of pieces in the kit, type of recording equipment, number of other instruments being recorded, as well as the type of sound desired [Shea 05].

The process of recording a piece of music involves the usage of a multitracker. A multitracker is software[25] or hardware based, having a certain number of tracks available for recording. Various tracks can be used to record only one instrument, but it is only possible to use one microphone per track. This is what happens with a drum kit. Using as an example the standard pop/rock drum kit (chapter 3.1), usually kick, snare, and hi-hat are recorded on individual tracks, as are each of the toms. However for the cymbals, overhead[26] microphones can be used to capture the sound of the instruments independently of the number of cymbals in the drum kit. Another very popular microphone setup implies the usage of the overhead microphones for the tapping of the toms also [Shea 05]. Figure 9.1 shows a drum kit ready for recording.

Figure 9.1 – A drum kit ready for recording.
Highlighted by red boxes are the overhead microphones.

25 Cubase from Steinberg, Pro Tools from Digidesign, and Sonar from Cakewalk are some of the examples of some of the most used software based multitrackers.

26 The name says it all, these microphones are placed above the drum kit. Usually two are used, one by the right side of the drum kit and another by the left side of the drum kit.

Although having dedicated microphones to almost every piece of the drum kit, each one of them captures the rest of the elements that are played but with a lesser level of amplitude then the assigned piece. When the drum kit recording is concluded, the whole collection of tracks is mixed to a single channel (mono), or into a two channel setup (stereo) [Shea 05], i.e., the sound source mixture.

. Attachment #2

This next attachment is comprised of the entire collection of tables with the values outputted by each one of the different tests we performed. Each line in any table corresponds to the samples of the particular cymbal to whom the line is connected to. If a line is colored in black then that particular test sample was classified inaccurately. T.P. (total points) is the total number of points from the sound sources of every test sample. A.C.P. (accurately classified points) is the number of points from T.P. that were correctly classified. S.R.(%) (success ratio) is the ratio between A.C.P. and T.P.. Avg(%) is the mean between the success ratios from all the test samples of a given cymbal. The last column of the tables has two different meanings. In the tests with two cymbals it gives us the number of samples that had more than 50% of accurately classified points. In the tests with three cymbals it gives us the number of samples whose majority of accurately classified points is bigger than the number of points badly classified that are distributed for each of the other two cymbals. Black lines are correspond to wrong classifications.

. Sets With High Amplitude Training Samples for Two Cymbals (Test #1)

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	227	205	90,308	91,3	6
	227	219	96,476		
	219	198	90,411		
	223	202	90,583		
	228	204	89,474		
	228	207	90,789		
China Edge	291	240	82,474	80,8	6
	288	245	85,069		
	286	227	79,371		
	267	226	84,644		
	132	106	80,303		
	134	98	73,134		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Crash 14 Edge	296	138	46,622	54,3	2
	274	90	32,847		
	300	144	48		
	234	113	48,291		
	227	176	77,533		
	187	136	72,727		
Crash 16 Edge	499	431	86,373	67,9	5
	499	398	79,76		
	499	345	69,138		
	499	347	69,539		
	499	320	64,128		
	249	96	38,554		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	227	209	92,07	64,4	4
	227	211	92,952		
	219	198	90,411		
	223	196	87,892		
	228	39	17,105		
	228	14	6,1404		
Crash 16 Edge	499	498	99,8	91,70267	6
	499	495	99,198		
	499	481	96,393		
	499	487	97,595		
	499	498	99,8		
	249	143	57,43		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
China Edge	291	201	69,072	75,9	6
	288	197	68,403		
	286	203	70,979		
	267	186	69,663		
	132	115	87,121		
	134	121	90,299		
Crash 16 Edge	499	471	94,389	84,9	5
	499	459	91,984		
	499	454	90,982		
	499	454	90,982		
	499	464	92,986		
	249	119	47,791		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
HH Cls Bow	249	249	100	100,0	6
	249	249	100		
	249	249	100		
	249	249	100		
	249	249	100		
	249	249	100		
Ride Bow	570	569	99,825	81,2	5
	474	465	98,101		
	434	423	97,465		
	478	465	97,28		
	424	222	52,358		
	396	168	42,424		

. Sets With Variable Amplitude Training Samples for Two Cymbals (Test #1)

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	218	217	99,541	98,8	6
	225	224	99,556		
	227	225	99,119		
	226	221	97,788		
	232	225	96,983		
	230	230	100		
China Edge	285	197	69,123	61,4	5
	291	200	68,729		
	272	189	69,485		
	271	190	70,111		
	221	116	52,489		
	132	51	38,636		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Crash 14 Edge	333	255	76,577	79,2	6
	297	281	94,613		
	258	188	72,868		
	261	191	73,18		
	230	176	76,522		
	200	163	81,5		
Crash 16 Edge	848	732	86,321	75,1	5
	499	453	90,782		
	499	424	84,97		
	499	407	81,563		
	499	402	80,561		
	249	66	26,506		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	218	209	95,872	73,5	5
	225	213	94,667		
	227	191	84,141		
	226	197	87,168		
	232	150	64,655		
	230	34	14,783		
Crash 16 Edge	848	832	98,113	99,55183	6
	499	499	100		
	499	498	99,8		
	499	330	99,8		
	499	499	100		
	249	248	99,598		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
China Edge	291	258	88,66	89,3	6
	288	275	95,486		
	286	190	66,434		
	267	228	85,393		
	132	132	100		
	134	134	100		
Crash 16 Edge	499	499	100	99,9	6
	499	499	100		
	499	499	100		
	499	498	99,8		
	499	499	100		
	249	248	99,598		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
HH Cls Bow	249	249	100	100,0	6
	249	249	100		
	249	249	100		
	249	249	100		
	249	249	100		
	249	249	100		
Ride Bow	589	589	100	87,1	6
	609	609	100		
	570	570	100		
	434	434	100		
	424	275	64,858		
	396	228	57,576		

. Sets With Low Amplitude Training Samples for Two Cymbals (Test #1)

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	190	190	78,421	93,2	6
	228	204	89,474		
	223	206	92,377		
	477	472	98,952		
	232	232	100		
	223	223	100		
China Edge	270	175	64,815	68,6	5
	285	171	60		
	285	251	88,07		
	293	160	54,608		
	267	201	75,281		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Crash 14 Edge	361	143	39,612	49,7	2
	333	160	48,048		
	297	188	63,3		
	274	107	39,051		
	267	127	47,566		
	227	137	60,352		
Crash 16 Edge	499	498	99,8	96,6	6
	499	499	100		
	499	499	100		
	499	473	94,79		
	499	458	91,784		
	499	465	93,186		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	190	185	97,368	87,2	5
	228	224	98,246		
	223	212	95,067		
	477	224	46,96		
	232	225	96,983		
	223	197	88,341		
Crash 16 Edge	499	499	100	100	6
	499	499	100		
	499	499	100		
	499	499	100		
	499	499	100		
	499	499	100		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
China Edge	270	268	99,259	93,2	5
	285	254	89,123		
	285	250	87,719		
	293	288	98,294		
	267	244	91,386		
Crash 16 Edge	499	499	100	99,9	6
	499	496	99,399		
	499	499	100		
	499	499	100		
	499	499	100		
	499	499	100		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
HH Cls Bow	249	249	100	99,8	5
	249	247	99,197		
	249	249	100		
	249	249	100		
	249	249	100		
Ride Bow	741	731	98,65	99,1	6
	589	579	98,302		
	614	607	98,86		
	511	507	99,217		
	570	569	99,825		
	474	473	99,789		

. Sets With High Amplitude Training Samples for Two Cymbals (Test #2)

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	52	52	100	96,7	6
	48	48	100		
	4	4	100		
	3	3	100		
	228	204	89,474		
	228	207	90,789		
China Edge	80	80	100	92,2	6
	80	80	100		
	63	63	100		
	63	63	100		
	132	106	80,303		
	134	98	73,134		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Crash 14 Edge	46	10	21,739	55,6	3
	38	17	44,737		
	19	3	15,789		
	1	1	100		
	227	180	79,295		
	187	135	72,193		
Crash 16 Edge	443	384	86,682	60,9	5
	238	201	84,454		
	199	142	71,357		
	140	86	61,429		
	86	53	61,628		
	1	0	0		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	43	42	97,674	66,1	4
	39	38	97,436		
	219	198	90,411		
	223	196	87,892		
	228	39	17,105		
	228	14	6,1404		
Crash 16 Edge	151	150	99,338	92,03483	6
	137	133	97,08		
	88	88	100		
	61	60	98,361		
	17	17	100		
	249	143	57,43		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
China Edge	47	15	31,915	46,5	2
	42	11	26,19		
	31	6	19,355		
	34	8	23,529		
	132	116	87,879		
	134	121	90,299		
Crash 16 Edge	136	114	83,824	81,5	5
	120	91	75,833		
	64	60	93,75		
	42	37	88,095		
	5	5	100		
	249	119	47,791		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
HH Cls Bow	8	8	100	100,0	6
	5	5	100		
	4	4	100		
	3	3	100		
	2	2	100		
	249	249	100		
Ride Bow	236	235	99,576	61,4	4
	147	138	93,878		
	91	80	87,912		
	78	68	87,179		
	2	0	0		
	1	0	0		

. Sets With Variable Amplitude Training Samples for Two Cymbals (Test #2)

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	57	57	100	99,5	6
	56	56	100		
	48	48	100		
	37	37	100		
	232	225	96,983		
	230	230	100		
China Edge	89	89	100	83,3	5
	93	93	100		
	89	89	100		
	80	80	100		
	21	21	100		
	1	0	0		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Crash 14 Edge	60	58	96,667	92,1	6
	60	57	95		
	54	52	96,296		
	58	52	89,655		
	10	10	100		
	200	150	75		
Crash 16 Edge	443	383	86,456	61,7	5
	238	203	85,294		
	199	145	72,864		
	139	87	62,59		
	86	54	62,791		
	1	0	0		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	54	54	100	80,0	5
	55	55	100		
	45	45	100		
	31	31	100		
	232	150	64,655		
	230	35	15,217		
Crash 16 Edge	376	375	99,734	99,95567	6
	204	204	100		
	183	183	100		
	126	126	100		
	83	83	100		
	6	6	100		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
China Edge	50	50	100	100,0	6
	50	50	100		
	33	33	100		
	32	32	100		
	132	132	100		
	134	134	100		
Crash 16 Edge	197	166	84,264	79,6	6
	178	153	85,955		
	133	105	78,947		
	121	95	78,512		
	61	61	100		
	2	1	50		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
HH Cls Bow	10	10	100	100,0	6
	9	9	100		
	7	7	100		
	4	4	100		
	3	3	100		
	249	249	100		
Ride Bow	398	398	100	66,7	4
	378	378	100		
	295	295	100		
	131	131	100		
	4	0	0		
	2	0	0		

. Sets With Low Amplitude Training Samples for Two Cymbals (Test #2)

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	176	134	77,841	92,8	6
	195	174	89,231		
	212	195	91,981		
	228	223	97,807		
	212	212	100		
	187	187	100		
China Edge	270	175	64,815	70,9	5
	255	171	67,059		
	270	251	92,963		
	293	160	54,608		
	267	201	75,281		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Crash 14 Edge	317	138	43,533	50,6	3
	328	162	49,39		
	245	132	53,878		
	274	117	42,701		
	267	136	50,936		
	227	143	62,996		
Crash 16 Edge	442	441	99,774	96,1	6
	451	451	100		
	379	379	100		
	358	330	91,62		
	362	331	91,436		
	340	318	93,529		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	111	106	95,495	98,5	6
	116	112	96,552		
	123	122	99,187		
	115	115	100		
	102	102	100		
	85	85	100		
Crash 16 Edge	442	442	100	100	6
	451	451	100		
	379	379	100		
	358	358	100		
	365	365	100		
	341	341	100		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
China Edge	183	183	100	100,0	5
	159	159	100		
	164	164	100		
	180	180	100		
	151	151	100		
Crash 16 Edge	442	442	100	99,9	6
	451	448	99,335		
	379	379	100		
	358	358	100		
	365	365	100		
	341	341	100		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
HH Cls Bow	18	18	100	91,8	5
	15	12	80		
	14	11	78,751		
	13	13	100		
	12	12	100		
Ride Bow	741	729	98,381	98,9	6
	589	578	98,132		
	614	605	98,534		
	511	506	99,022		
	570	569	99,825		
	474	473	99,789		

. Sets With High Amplitude Training Samples for Two Cymbals (Test #3)

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	99	99	100	96,7	6
	93	93	100		
	54	54	100		
	52	52	100		
	228	204	89,474		
	228	207	90,789		
China Edge	176	176	100	97,5	6
	175	175	100		
	159	159	100		
	159	159	100		
	57	52	91,228		
	50	47	94		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Crash 14 Edge	210	87	41,429	43,8	2
	215	90	41,86		
	199	82	41,206		
	162	88	54,321		
	76	61	80,263		
	25	1	4		
Crash 16 Edge	364	295	81,004	67,3	5
	351	253	72,08		
	342	196	57,31		
	323	172	53,251		
	308	136	44,156		
	26	25	96,154		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	87	86	98,851	70,3	4
	76	75	98,684		
	40	40	100		
	33	33	100		
	228	41	17,982		
	228	14	6,1404		
Crash 16 Edge	364	363	99,725	82,39613	5
	352	348	98,864		
	345	327	94,783		
	329	318	96,657		
	315	315	100		
	46	2	4,3478		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
China Edge	133	65	48,872	65,5	4
	135	66	48,889		
	126	67	53,175		
	123	66	53,659		
	20	19	95		
	15	14	93,333		
Crash 16 Edge	364	344	94,505	77,9	5
	349	308	88,252		
	341	297	87,097		
	321	279	86,916		
	305	270	88,525		
	18	4	22,222		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
HH Cls Bow	12	12	100	100,0	6
	10	10	100		
	9	9	100		
	6	6	100		
	6	6	100		
	6	6	100		
Ride Bow	570	569	99,825	96,9	6
	473	464	98,097		
	432	421	97,454		
	467	457	97,859		
	215	204	94,884		
	164	153	93,293		

. Sets With Variable Amplitude Training Samples for Two Cymbals (Test #3)

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	125	125	100	100,0	6
	122	122	100		
	109	109	100		
	101	101	100		
	63	63	100		
	1	1	100		
China Edge	191	191	100	94,5	6
	192	191	99,479		
	186	186	100		
	186	186	100		
	111	110	99,099		
	73	50	68,493		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Crash 14 Edge	303	204	67,327	68,4	6
	245	204	83,265		
	258	170	65,891		
	261	154	59,004		
	229	142	62,009		
	184	134	72,826		
Crash 16 Edge	848	760	89,623	71,7	5
	379	342	90,237		
	358	296	82,682		
	331	253	76,435		
	342	271	79,24		
	131	16	12,214		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	97	97	100	85,8	5
	98	98	100		
	87	87	100		
	84	84	100		
	28	28	100		
	230	34	14,783		
Crash 16 Edge	848	829	97,759	99,6265	6
	379	379	100		
	358	358	100		
	331	331	100		
	341	341	100		
	122	122	100		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
China Edge	134	134	100	100,0	6
	136	136	100		
	136	136	100		
	118	118	100		
	37	37	100		
	23	23	100		
Crash 16 Edge	364	364	100	100,0	6
	356	356	100		
	355	355	100		
	331	331	100		
	325	325	100		
	94	94	100		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
HH Cls Bow	15	15	100	100,0	6
	14	14	100		
	12	12	100		
	8	8	100		
	7	7	100		
	6	6	100		
Ride Bow	589	589	100	98,5	6
	609	609	100		
	570	570	100		
	434	434	100		
	268	257	95,896		
	226	215	95,133		

. **Sets With Low Amplitude Training Samples for Two Cymbals (Test #3)**

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	190	149	78,421	93,0	6
	228	204	89,474		
	213	196	92,019		
	229	224	97,817		
	225	225	100		
	197	197	100		
China Edge	270	175	64,815	70,6	5
	259	171	66,023		
	272	251	92,279		
	293	160	54,608		
	267	201	75,281		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Crash 14 Edge	358	153	42,737	49,9	3
	333	168	50,45		
	297	145	48,822		
	274	119	43,431		
	267	135	50,562		
	227	144	63,436		
Crash 16 Edge	442	441	99,774	96,1	6
	451	451	100		
	379	379	100		
	359	329	91,643		
	366	336	91,803		
	343	321	93,586		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
Splash Edge	190	185	97,368	98,8	6
	228	224	98,246		
	213	212	99,531		
	229	224	97,817		
	225	225	100		
	197	197	100		
Crash 16 Edge	442	442	100	100	6
	451	451	100		
	379	379	100		
	359	359	100		
	366	366	100		
	343	343	100		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
China Edge	270	268	99,259	96,2	5
	255	254	99,608		
	272	251	92,279		
	293	288	98,294		
	267	244	91,386		
Crash 16 Edge	442	442	100	99,9	6
	451	448	99,335		
	379	379	100		
	359	359	100		
	366	366	100		
	343	343	100		

	T.P.	A.C.P.	S.R.(%)	Avg(%)	>= 50%
HH Cls Bow	249	249	100	99,8	5
	249	247	99,197		
	249	249	100		
	249	249	100		
	249	249	100		
Ride Bow	741	731	98,65	99,1	6
	589	580	98,472		
	614	607	98,86		
	511	507	99,217		
	570	569	99,825		
	474	473	99,789		

. **Sets With High Amplitude Training Samples for Three Cymbals (Test #1)**

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. China	S.R.(%)	Avg(%)	
Splash Edge	227	165	61	1	72,687	46,3	4
	227	163	63	1	71,806		
	219	138	81	0	63,014		
	223	121	102	0	54,26		
	228	30	184	14	13,158		
	228	6	210	12	2,6316		
			A.C.P. Splash	A.C.P. China			
Crash 16 Edge	499	499	0	0	100	94,7	6
	499	713	1	0	99,8		
	499	499	0	0	100		
	499	498	1	0	99,8		
	499	499	0	0	100		
	249	171	78	0	68,675		
			A.C.P. Splash	A.C.P. Crash 16			
China Edge	291	49	175	67	16,838	11,6	0
	288	39	195	54	13,542		
	286	55	142	89	19,231		
	267	53	147	67	19,85		
	132	0	129	3	0		
	134	0	106	28	0		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
Splash Edge	227	131	40	56	57,709	36,0	3
	227	147	49	31	64,758		
	219	93	59	67	42,466		
	223	74	60	89	33,184		
	228	31	197	0	13,596		
	228	10	210	0	4,386		
			A.C.P. Splash	A.C.P. Crash 14			
Crash 16 Edge	499	499	0	0	100	97,6	6
	499	498	1	0	99,8		
	499	499	0	0	100		
	499	498	1	0	99,8		
	499	499	0	0	100		
	249	214	27	8	85,944		
			A.C.P. Splash	A.C.P. Crash 16			
Crash 14 Edge	296	60	18	218	20,27	8,7	0
	274	52	219	3	18,978		
	300	24	0	276	8		
	234	0	234	0	0		
	227	11	8	208	4,8458		
	187	0	69	118	0		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
China Edge	45	45	0	0	100	100,0	6
	44	44	0	0	100		
	51	51	0	0	100		
	28	28	0	0	100		
	5	5	0	0	100		
	3	3	0	0	100		
			A.C.P. China	A.C.P. Crash 14			
Crash 16 Edge	364	363	0	1	99,725	98,5	6
	350	349	0	1	99,714		
	346	339	0	7	97,977		
	328	328	0	0	100		
	315	312	0	2	99,365		
	51	48	1	2	94,118		
			A.C.P. China	A.C.P. Crash 16			
Crash 14 Edge	174	169	2	3	97,126	81,2	5
	189	168	0	21	88,889		
	161	158	0	3	98,137		
	156	65	91	0	41,667		
	52	44	8	0	84,615		
	17	13	4	0	76,471		

. Sets With Variable Amplitude Training Samples for Three Cymbals (Test #1)

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. China	S.R.(%)	Avg(%)	
Splash Edge	218	205	9	4	94,037	73,8	5
	225	202	20	3	89,778		
	227	176	48	3	77,533		
	226	190	35	1	84,071		
	232	142	89	1	61,207		
	230	83	143	4	36,087		
			A.C.P. Splash	A.C.P. China			
Crash 16 Edge	848	42	276	530	4,9528	81,9	5
	499	486	0	13	97,395		
	499	495	1	3	99,198		
	499	498	1	0	99,8		
	499	498	0	1	99,8		
	249	224	25	0	89,96		
			A.C.P. Splash	A.C.P. Crash 16			
China Edge	285	65	220	0	22,807	14,0	0
	291	57	234	0	19,588		
	272	54	218	0	19,853		
	271	41	230	0	15,129		
	221	6	215	0	2,7149		
	132	5	127	0	3,7879		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	>= 50%
Splash Edge	218	197	13	8	90,367	70,6	5
	225	192	21	12	85,333		
	227	169	49	9	74,449		
	226	175	35	16	77,434		
	232	138	89	5	59,483		
	230	**84**	**146**	**0**	**36,522**		
			A.C.P. Splash	A.C.P. Crash 14			
Crash 16 Edge	**848**	**128**	**580**	**140**	**15,094**	78,2	5
	499	464	6	29	92,986		
	499	476	1	22	95,391		
	499	446	6	47	89,379		
	499	462	5	32	92,585		
	249	209	0	40	83,936		
			A.C.P. Splash	A.C.P. Crash 16			
Crash 14 Edge	333	126	94	113	37,838	40,0	4
	297	127	91	79	42,761		
	258	111	38	109	43,023		
	261	141	42	78	54,023		
	230	**99**	**20**	**111**	**43,043**		
	200	**39**	**16**	**145**	**19,5**		

f

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
China Edge	148	146	0	2	98,649	99,8	6
	128	128	0	0	100		
	97	97	0	0	100		
	124	124	0	0	100		
	56	56	0	0	100		
	5	5	0	0	100		
			A.C.P. China	A.C.P. Crash 14			
Crash 16 Edge	848	587	86	175	69,222	94,5	6
	379	379	0	0	100		
	358	353	0	5	98,603		
	331	327	0	4	98,92		
	341	341	0	0	100		
	97	97	0	0	100		
			A.C.P. China	A.C.P. Crash 16			
Crash 14 Edge	265	171	6	88	64,528	56,8	4
	179	127	33	19	70,95		
	213	149	0	64	69,953		
	256	**127**	**0**	**129**	**49,609**		
	169	101	0	68	59,763		
	157	**41**	**0**	**116**	**26,115**		

. Sets With Low Amplitude Training Samples for Three Cymbals (Test #1)

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. China	S.R.(%)	Avg(%)	
Splash Edge	157	103	3	51	65,605	72,5	6
	172	125	7	40	72,674		
	208	146	22	40	70,192		
	229	173	17	39	75,546		
	224	160	51	13	71,429		
	196	156	38	2	79,592		
			A.C.P. Splash	A.C.P. China			
Crash 16 Edge	442	442	0	0	100	100,0	6
	451	451	0	0	100		
	379	379	0	0	100		
	359	359	0	0	100		
	366	366	0	0	100		
	343	342	1	0	99,708		
			A.C.P. Splash	A.C.P. Crash 16			
China Edge	270	251	19	0	92,963	92,9	5
	255	255	0	0	100		
	272	270	1	1	99,265		
	293	251	41	1	85,666		
	267	231	32	4	86,517		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
Splash Edge	190	167	8	15	87,895	94,7	6
	213	194	6	13	91,08		
	205	191	2	12	93,171		
	191	183	0	8	95,812		
	177	177	0	0	100		
	170	170	0	0	100		
			A.C.P. Splash	A.C.P. Crash 14			
Crash 16 Edge	442	442	0	0	100	98,4	6
	451	451	0	0	100		
	379	379	0	0	100		
	359	349	0	10	97,214		
	365	355	0	10	97,26		
	343	329	0	14	95,918		
			A.C.P. Splash	A.C.P. Crash 16			
Crash 14 Edge	356	149	52	155	41,854	53,3	4
	333	162	33	138	48,649		
	297	180	60	57	60,606		
	274	121	27	126	44,161		
	267	133	22	112	49,813		
	227	169	9	49	74,449		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	>= 50%
China Edge	270	263	7	0	97,407	92,4	5
	254	254	0	0	100		
	267	257	10	0	96,255		
	293	232	61	0	79,181		
	262	233	29	0	88,931		
			A.C.P. China	A.C.P. Crash 14			
Crash 16 Edge	210	210	0	0	100	98,6	6
	260	257	0	3	98,846		
	238	238	0	0	100		
	259	256	0	3	98,842		
	297	295	0	2	99,327		
	210	199	0	11	94,762		
			A.C.P. China	A.C.P. Crash 16			
Crash 14 Edge	**212**	**71**	**135**	**6**	**33,491**	60,4	3
	209	**96**	**107**	**6**	**45,933**		
	209	**78**	**131**	**0**	**37,321**		
	174	139	35	0	79,885		
	179	148	30	1	82,682		
	152	126	26	0	82,895		

. Sets With High Amplitude Training Samples for Three Cymbals (Test #2)

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. China	S.R.(%)	Avg(%)	
Splash Edge	36	35	0	1	97,222	54,6	4
	36	35	0	1	97,222		
	219	136	86	0	62,1		
	223	123	100	0	55,157		
	228	**30**	**184**	**14**	**13,158**		
	228	**6**	**210**	**12**	**2,6316**		
			A.C.P. Splash	A.C.P. China			
Crash 16 Edge	101	101	0	0	100	93,7	6
	91	91	0	0	100		
	81	81	0	0	100		
	57	57	0	0	100		
	20	20	0	0	100		
	249	155	94	0	62,249		
			A.C.P. Splash	A.C.P. Crash 16			
China Edge	38	38	0	0	100	63,7	4
	34	28	6	0	82,353		
	13	13	0	0	100		
	12	12	0	0	100		
	132	**0**	**130**	**2**	**0**		
	134	**0**	**112**	**22**	**0**		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
Splash Edge	227	138	30	59	60,793	36,7	4
	227	142	59	26	62,555		
	219	101	59	59	46,119		
	223	82	69	72	36,771		
	228	23	205	0	10,088		
	228	9	219	0	3,9474		
			A.C.P. Splash	A.C.P. Crash 14			
Crash 16 Edge	499	499	0	0	100	91,9	6
	499	498	0	1	99,8		
	499	499	0	0	100		
	499	498	0	1	99,8		
	499	498	1	0	99,8		
	249	130	0	119	52,209		
			A.C.P. Splash	A.C.P. Crash 16			
Crash 14 Edge	296	51	3	242	17,23	14,8	0
	274	53	0	221	19,343		
	300	17	0	283	5,667		
	234	0	0	234	0		
	227	27	18	182	11,894		
	187	65	15	107	34,759		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
China Edge	125	80	0	45	64	73,9	6
	124	81	0	43	65,323		
	102	65	0	37	63,725		
	104	70	0	34	67,308		
	18	15	0	3	83,333		
	11	11	0	0	100		
			A.C.P. China	A.C.P. Crash 14			
Crash 16 Edge	360	360	0	0	100	96,1	6
	349	349	0	0	100		
	336	336	0	0	100		
	314	314	0	0	100		
	300	300	0	0	100		
	249	191	0	58	76,707		
			A.C.P. China	A.C.P. Crash 16			
Crash 14 Edge	201	80	0	121	39,801	15,8	0
	194	72	0	122	37,113		
	175	30	0	145	17,143		
	158	1	0	157	0,63291		
	66	0	66	0	0		
	1	0	1	0	0		

. Sets With Variable Amplitude Training Samples for Three Cymbals (Test #2)

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. China	S.R.(%)	Avg(%)	
Splash Edge	43	43	0	0	100	78,0	5
	43	42	0	1	97,674		
	34	33	0	1	97,059		
	15	15	0	0	100		
	232	125	89	18	53,879		
	230	45	185	0	19,565		
			A.C.P. Splash	A.C.P. China			
Crash 16 Edge	324	166	10	148	51,235	91,9	6
	186	186	0	0	100		
	169	169	0	0	100		
	101	101	0	0	100		
	65	65	0	0	100		
	1	1	0	0	100		
			A.C.P. Splash	A.C.P. Crash 16			
China Edge	31	29	2	0	93,548	78,1	4
	28	28	0	0	100		
	22	22	0	0	100		
	8	8	0	0	100		
	221	91	112	18	41,176		
	132	45	87	0	34,091		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
Splash Edge	218	190	11	17	87,156	70,1	5
	225	194	19	12	86,222		
	227	165	51	11	72,687		
	226	182	31	13	80,531		
	232	134	89	9	57,759		
	230	83	147	0	36,087		
			A.C.P. Splash	A.C.P. Crash 14			
Crash 16 Edge	848	116	575	157	13,679	77,1	5
	499	446	5	48	89,379		
	499	472	0	27	94,589		
	499	440	5	54	88,176		
	499	460	4	35	92,184		
	249	210	0	39	84,337		
			A.C.P. Splash	A.C.P. Crash 16			
Crash 14 Edge	333	129	95	109	38,739	42,7	4
	297	126	91	80	42,424		
	258	120	35	103	46,512		
	261	147	53	61	56,322		
	230	101	25	104	43,913		
	200	126	17	57	28,5		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
China Edge	147	146	0	1	99,32	99,9	6
	152	152	0	0	100		
	137	137	0	0	100		
	135	135	0	0	100		
	70	70	0	0	100		
	41	41	0	0	100		
			A.C.P. China	A.C.P. Crash 14			
Crash 16 Edge	848	592	86	170	69,811	94,6	6
	379	379	0	0	100		
	358	354	0	4	98,883		
	331	327	0	4	98,792		
	341	341	0	0	100		
	96	96	0	0	100		
			A.C.P. China	A.C.P. Crash 16			
Crash 14 Edge	265	173	4	88	65,283	56,5	4
	180	131	29	20	72,778		
	211	145	0	66	68,72		
	254	**123**	**0**	**131**	**48,425**		
	165	99	0	66	60		
	154	**37**	**0**	**117**	**24,026**		

. Sets With Low Amplitude Training Samples for Three Cymbals (Test #2)

	T.P.	A.C.P.	A.C.P. Crash 16	.C.P. Chin	S.R.(%)	Avg(%)	
Splash Edge	155	103	2	50	66,452	76,0	6
	167	122	6	39	73,054		
	208	158	10	40	75,962		
	229	182	9	38	79,476		
	224	175	37	12	78,125		
	196	162	32	2	82,653		
			A.C.P. Splash	.C.P. Chin			
Crash 16 Edge	442	442	0	0	100	100,0	6
	451	451	0	0	100		
	379	379	0	0	100		
	359	359	0	0	100		
	366	366	0	0	100		
	343	343	0	0	99,708		
			A.C.P. Splash	C.P. Crash			
China Edge	270	251	19	0	92,963	93,0	5
	255	255	0	0	100		
	272	271	0	1	99,632		
	293	251	41	1	85,666		
	267	231	32	4	86,517		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
Splash Edge	190	168	6	16	88,421	95,1	6
	209	192	5	12	91,866		
	207	196	2	9	94,686		
	192	184	0	8	95,833		
	177	177	0	0	100		
	169	169	0	0	100		
			A.C.P. Splash	A.C.P. Crash 14			
Crash 16 Edge	442	442	0	0	100	98,4	6
	451	451	0	0	100		
	379	379	0	0	100		
	359	349	0	10	97,214		
	365	356	0	9	97,534		
	343	329	0	14	95,918		
			A.C.P. Splash	A.C.P. Crash 16			
Crash 14 Edge	357	153	49	155	42,857	53,5	4
	333	161	32	140	48,348		
	297	194	50	53	65,32		
	274	118	28	128	43,006		
	267	127	24	116	47,566		
	227	168	11	48	74,009		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
China Edge	270	238	32	0	88,148	84,2	5
	255	246	9	0	96,471		
	272	248	24	0	91,176		
	293	214	79	0	73,038		
	267	193	74	0	72,285		
			A.C.P. China	A.C.P. Crash 14			
Crash 16 Edge	442	442	0	0	100	99,0	6
	451	446	0	5	98,891		
	379	379	0	0	100		
	359	354	0	5	98,607		
	364	361	1	2	99,176		
	343	333	1	9	97,085		
			A.C.P. China	A.C.P. Crash 16			
Crash 14 Edge	357	179	148	30	50,14	69,8	5
	333	192	128	13	57,658		
	297	130	157	10	43,771		
	274	245	29	0	89,416		
	267	239	26	2	89,513		
	227	200	27	0	88,106		

. Sets With High Amplitude Training Samples for Three Cymbals (Test #3)

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. China	S.R.(%)	Avg(%)	
Splash Edge	78	77	0	1	98,718	68,9	4
	72	71	0	1	98,611		
	30	30	0	0	100		
	21	21	0	0	100		
	228	30	184	14	13,158		
	228	6	210	12	2,6316		
			A.C.P. Splash	A.C.P. China			
Crash 16 Edge	360	360	0	0	100	94,2	6
	332	332	0	0	100		
	301	301	0	0	100		
	288	288	0	0	100		
	258	258	0	0	100		
	249	162	87	0	65,06		
			A.C.P. Splash	A.C.P. Crash 16			
China Edge	129	48	79	2	37,209	27,5	1
	127	36	88	3	28,346		
	105	55	47	3	52,381		
	111	52	55	4	46,847		
	21	0	21	0	0		
	15	0	15	0	0		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
Splash Edge	227	165	53	9	72,687	44,0	4
	227	147	63	17	64,758		
	219	117	77	25	53,425		
	223	102	92	29	45,74		
	228	44	184	0	19,298		
	228	19	209	0	8,3333		
			A.C.P. Splash	A.C.P. Crash 14			
Crash 16 Edge	499	499	0	0	100	93,6	6
	499	498	0	1	99,8		
	499	499	0	0	100		
	499	498	0	1	99,8		
	499	498	1	0	99,8		
	249	155	0	94	62,249		
			A.C.P. Splash	A.C.P. Crash 16			
Crash 14 Edge	296	50	2	244	16,892	14,5	0
	274	51	0	223	18,613		
	300	17	0	283	5,6667		
	234	0	0	234	0		
	227	27	18	182	11,894		
	187	64	14	109	34,225		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
China Edge	123	80	0	43	65,041	64,7	5
	120	73	0	47	60,833		
	101	55	0	46	54,455		
	103	61	0	42	59,233		
	8	8	0	0	100		
	134	65	0	69	48,507		
			A.C.P. China	A.C.P. Crash 14			
Crash 16 Edge	364	364	0	0	100	100,0	6
	349	349	0	0	100		
	339	339	0	0	100		
	316	316	0	0	100		
	304	304	0	0	100		
	36	36	0	0	100		
			A.C.P. China	A.C.P. Crash 16			
Crash 14 Edge	205	80	0	125	39,024	15,3	0
	198	71	0	127	35,859		
	182	30	0	152	16,484		
	158	1	0	157	0,63291		
	69	0	0	69	0		
	9	0	0	9	0		

. Sets With Variable Amplitude Training Samples for Three Cymbals (Test #3)

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. China	S.R.(%)	Avg(%)	
Splash Edge	101	99	0	2	98,02	88,7	5
	91	90	0	1	98,901		
	86	84	0	2	97,674		
	82	82	0	0	100		
	39	39	0	0	100		
	230	86	140	4	37,391		
			A.C.P. Splash	A.C.P. China			
Crash 16 Edge	848	41	275	532	4,8349	83,4	5
	379	366	0	13	96,57		
	358	355	0	3	99,162		
	331	331	0	0	100		
	340	339	0	1	99,706		
	76	76	0	0	100		
			A.C.P. Splash	A.C.P. Crash 16			
China Edge	180	24	156	0	13,333	8,6	0
	176	26	147	3	14,773		
	162	19	143	0	11,728		
	145	17	128	0	11,724		
	88	0	88	0	0		
	42	0	42	0	0		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
Splash Edge	90	90	0	0	100	89,2	5
	86	86	0	0	100		
	71	71	0	0	100		
	80	80	0	0	100		
	31	31	0	0	100		
	230	81	149	0	35,217		
			A.C.P. Splash	A.C.P. Crash 14			
Crash 16 Edge	848	114	574	114	13,443	76,1	5
	379	310	5	64	81,794		
	358	330	0	28	92,179		
	331	270	5	56	81,571		
	341	298	4	39	87,39		
	98	98	0	0	100		
			A.C.P. Splash	A.C.P. Crash 16			
Crash 14 Edge	290	112	95	83	38,621	42,2	4
	234	83	90	61	35,47		
	258	118	38	102	45,736		
	261	146	53	62	55,939		
	204	96	25	83	47,059		
	175	53	17	105	30,286		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
China Edge	123	123	0	0	100	100,0	6
	127	127	0	0	100		
	123	123	0	0	100		
	121	121	0	0	100		
	58	58	0	0	100		
	37	37	0	0	100		
			A.C.P. China	A.C.P. Crash 14			
Crash 16 Edge	311	272	0	0	1,2862	83,5	5
	138	138	0	0	100		
	149	149	0	0	100		
	85	85	0	0	100		
	75	75	0	0	100		
	17	17	0	0	100		
			A.C.P. China	A.C.P. Crash 16			
Crash 14 Edge	114	77	37	0	67,544	72,1	6
	112	71	41	0	63,393		
	90	84	6	0	93,333		
	90	75	15	0	83,333		
	29	23	1	5	79,31		
	11	5	1	5	45,455		

. Sets With Low Amplitude Training Samples for Three Cymbals (Test #3)

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. China	S.R.(%)	Avg(%)	
Splash Edge	179	117	8	54	65,363	76,8	6
	201	129	29	43	64,179		
	169	127	1	41	75,148		
	142	103	0	39	72,535		
	132	116	0	16	87,879		
	131	125	3	3	95,42		
			A.C.P. Splash	A.C.P. China			
Crash 16 Edge	442	442	0	0	100	99,9	6
	451	451	0	0	100		
	379	379	0	0	100		
	359	359	0	0	100		
	366	366	0	0	100		
	343	341	2	0	99,417		
			A.C.P. Splash	A.C.P. Crash 16			
China Edge	270	251	19	0	92,963	92,4	5
	255	255	0	0	100		
	272	270	1	1	99,265		
	293	245	47	1	83,618		
	267	230	34	3	86,142		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
Splash Edge	151	127	6	18	84,106	93,8	6
	151	133	5	13	88,079		
	208	196	2	10	94,231		
	229	221	0	8	96,507		
	224	224	0	0	100		
	195	195	0	0	100		
			A.C.P. Splash	A.C.P. Crash 14			
Crash 16 Edge	442	442	0	0	100	98,5	6
	451	451	0	0	100		
	379	379	0	0	100		
	359	349	0	10	97,214		
	366	357	0	9	97,541		
	343	331	0	12	96,501		
			A.C.P. Splash	A.C.P. Crash 16			
Crash 14 Edge	357	151	50	156	42,297	51,6	3
	333	153	33	147	45,946		
	297	197	50	50	66,33		
	274	111	29	134	40,511		
	267	118	26	123	44,195		
	227	160	12	55	70,485		

	T.P.	A.C.P.	A.C.P. Crash 16	A.C.P. Crash 14	S.R.(%)	Avg(%)	
Splash Edge	270	261	9	0	**96,667**	92,6	5
	255	255	0	0	**100**		
	272	265	7	0	**97,426**		
	293	240	53	0	**81,911**		
	267	232	35	0	**86,891**		
			A.C.P. China	**A.C.P. Crash 14**			
Crash 16 Edge	442	441	0	1	**99,774**	98,8	6
	451	448	0	3	**99,335**		
	379	379	0	0	**100**		
	356	353	0	3	**99,157**		
	362	360	0	2	**99,448**		
	338	322	1	15	**95,266**		
			A.C.P. China	**A.C.P. Crash 16**			
Crash 14 Edge	357	179	145	33	**50,14**	71,0	5
	333	213	112	8	**63,964**		
	297	**141**	**156**	**0**	**47,475**		
	274	237	36	1	**86,496**		
	267	237	29	1	**88,764**		
	227	203	24	0	**89,427**		