UNIVERSITE PARIS VI - PIERRE ET MARIE CURIE

ECOLE DOCTORALE D'INFORMATIQUE
TELECOMMUNICATIONS ET ELECTRONIQUE (EDITE PARIS)

These de doctorat

Discipline : Informatique

# MULTIOBJECTIVE TIME SERIES MATCHING AND CLASSIFICATION

ESLING PHILIPPE

Under supervision of AGON Carlos

Defended on December 5th 2012

In front of the following jury

| | | |
|---|---|---|
| **AMMARI Habib** | Directeur de recherche (CNRS) | Rapporteur |
| **SLANEY Malcolm** | Professor (Stanford university) | Rapporteur |
| **AGON Carlos** | Professeur des universités (UPMC) | Directeur |
| **ALLOUCHE Jean-Paul** | Directeur de recherche (CNRS) | Examinateur |
| **PACHET François** | Senior researcher (Sony CSL) | Examinateur |
| **PERNY Patrice** | Professeur des universités (UPMC) | Examinateur |
| **PRESSNITZER Daniel** | Directeur de recherche (CNRS) | Examinateur |

Equipe représentations musicales

Institut de Recherche et Coordination Acoustique Musique (IRCAM)

*"Education is an admirable thing, but it is well to remember from time to time that nothing that is worth knowing can be taught."*

— Oscar Wilde

# ABSTRACT


Millions of years of genetic evolution have shaped our auditory system, raising our way of listening to a form of art. Despite a somehow limited frequency spectrum, we are able to achieve an excellent and flexible discrimination of events. These unique capacities originate from the ability of our brain to organize our perception of sounds and music. We can process several conflicting scales simultaneously, thus constructing a multidimensional structure of perception. Furthermore, even if time is an ubiquitous and complex concept, humans have a natural capacity to extract meaningful knowledge from the shape of temporal structures. The onset of this study was, therefore, to explore these temporal and perceptual aspects in order to create a framework for generating musical orchestrations.

We show that by drawing inspiration from our musical perception and gaining insights from these mechanisms to drive our choices of algorithms, we can create innovative and powerful approaches for generic querying and classification, far outside the realm of musical problematics. First, by trying to emulate our multiobjective perception of temporal structures, we propose a framework called *MultiObjective Time Series* (MOTS) matching. We formally state this novel problem and provide an efficient algorithm to solve it. Based on this approach, we are able to introduce two innovative audio querying paradigms. We propose an experimental protocol to examine their effectiveness and usability through user studies. We analyze the validity of our proposal by studying the perception of the temporal evolution of conflicting higher-level audio features. We reveal the concept of multidimensional *directions of listening* that forms in the brain. We show that these directions are consistent through various tasks and unique to each person. We further propose a novel and flexible classification model based on the *hypervolumes dominated* by different classes, called *HyperVolume-MOTS* (HV-MOTS) classification. Instead of trying to consider the position of an element with respect to the various classes, this framework studies the behavior of each class with respect to the input through the distribution and spread over the optimization space. We show that the multiobjective flexibility inspired by our musical perception produces a classification paradigm that outperforms state-of-the-art methods on a wide range of scientific problems such as EEG analysis, climatology, medical diagnosis, character recognition and robotics. We present a comparison of this classification paradigm to traditional classifiers such as Nearest-Neighbor, Nearest-Center or Support Vector Machines. We then perform a thorough evaluation of our new approach and demonstrate its superiority on a wide range of datasets. We also show several applications of this scheme and study its weaknesses and strengths in each case. We present our main finding in which this method allows to construct a biometric identification system based on the sounds produced by heartbeats. We specifically develop for this problem a novel set of features based on the Stockwell transform and inspired by research in musical analysis. We show that we can accurately identify human beings through the sounds their heart produce. Our system obtains error rates equivalent to other biometrics such as face or speech recognition. These findings are supported by the largest heart sounds dataset ever collected, including the Mars500 isolation study.

Finally, we show how all this knowledge gained allows to come back to our initial artistic problematics of musical orchestration. We consider the problem of generating orchestral sound mixtures that can closely approximate any given audio signal. While

performing this reconstruction, we avoid mixing the similarities into a single measure, but rather use an advanced search algorithm based on the MOTS framework, called optimal warping. This allows us to obtain a set of efficient solutions that provide various compromises among spectral objectives. This algorithm performs a morphological segmentation procedure based on the variation of entropy. We then present several musical applications and interfaces that result from our various generic findings.

# RÉSUMÉ

Plusieurs millions d'années d'évolution génétique ont façonné notre système auditif, élevant ainsi notre écoute au rang d'un art. Malgré un spectre de fréquences perçues quelque peu limité, nous sommes en mesure d'effectuer une discrimination précise et flexible des événements auditifs. Ces capacités uniques proviennent de la capacité qu'a notre cerveau à organiser notre perception des sons et de la musique. Nous pouvons ainsi traiter simultanément plusieurs échelles de perception contradictoires, par la construction d'une structure multidimensionnelle de la perception. De plus, même si le temps est un concept omniprésent et complexe, les êtres humains ont une capacité inhérente à extraire une structure cohérente à partir de formes temporelles. Le point de départ de notre travail était donc d'étudier ces aspects temporels et perceptuels pour la création d'un système de génération d'orchestration musicale.

Nous montrons qu'en s'inspirant de cette perception musicale et en émulant ces mécanismes dans nos choix algorithmiques, nous sommes en mesure de créer des approches novatrices et efficaces de recherche et de classification générique, dépassant largement le cadre des problématiques musicales. Tout d'abord, en essayant d'imiter le caractère multi-objectif de notre perception des structures temporelles, nous proposons un cadre de recherche appelé *MultiObjective Time Series* (MOTS). Nous commençons par définir formellement ce nouveau problème et proposons un algorithme efficace pour le résoudre. Sur la base de cette approche, nous sommes en mesure d'introduire deux paradigmes innovants de recherche sur les fichiers audio. Nous étudions l'efficacité et la facilité d'utilisation de ces paradigmes grâce à des études utilisateurs. Grâce à cette étude, nous analysons également la validité de notre proposition en analysant la perception d'évolutions temporelles conflictuelles sur des descripteurs audio de haut niveau. Nous exposons ainsi le concept de *directions d'écoute* multidimensionnelles qui prends naissance dans notre perception. Nous montrons que ces directions sont consistantes à travers plusieurs tâches mais également uniques à chaque personne. Après cette validation, nous introduisons un nouveau paradigme flexible de classification basé sur les *hypervolumes dominés* par les différentes classes, appelé *HyperVolume-MOTS* (HV-MOTS). Contrairement aux paradigmes classiques qui étudient la position d'un élément par rapport aux différentes classes existantes, notre système étudie le comportement de la classe entière par rapport à l'élement à travers la distribution et la diffusion d'une classe sur l'espace d'optimisation. Nous montrons que la flexibilité multi-objective inspirée par notre perception musicale produit un paradigme de classification qui surpasse les méthodes de l'état de l'art sur un large éventail de problèmes scientifiques tels que l'analyse EEG, la climatologie, le diagnostic médical, la reconnaissance de caractères et la robotique. Nous fournissons une comparaison de ce paradigme par rapport aux classificateurs classiques tels que le Nearest-Neighbor, Nearest-Center ou Support Vector Machines. Nous effectuons ensuite une évaluation exhaustive et approfondie de notre nouvelle approche et démontrons sa supériorité sur un large ensemble de données. Nous montrons en outre plusieurs applications permettant d'étudier de manière plus détaillée les forces et faiblesses de notre proposition. Nous présentons l'application principale de cette méthode dans laquelle elle permet de construire un système d'identification biométrique basée sur les sons produit par les battements de coeur. En particulier, nous développons pour ce problème un nouvel ensemble de descripteurs basés sur la transformée de Stockwell et inspiré par la recherche en analyse

musicale. Nous montrons que nous pouvons identifier avec précision les êtres humains à travers les sons que produit leur cœur et que nous atteignons des taux d'erreur équivalent à d'autres caractéristiques biométriques telles que la reconnaissance vocale. Ces résultats sont confirmés par le plus grand ensemble de données de sons cardiaques jamais recueillies, comprenant également l'étude d'isolation Mars500 effectuée par l'Agence Spatiale Européenne.

Enfin, nous montrons comment toute cette connaissance acquise permet de revenir à nos problématiques artistiques originales d'orchestration musicale. Nous étudions ainsi le problème de la génération de mélanges sonores orchestraux imitant au mieux une cible audio donnée. En effectuant cette reconstruction, nous évitons de mélanger la similarité en une mesure de distance unique et nous utilisons un nouvel algorithme de recherche basé sur le cadre MOTS appelé *Optimal Warping*. Cette approche nous permet ainsi d'obtenir un ensemble de solutions efficaces qui offrent différents compromis entre les objectifs spectraux. Cet algorithme effectue une segmentation morphologique basée sur l'analyse de la variation d'entropie des séries temporelles. Nous présentons enfin plusieurs interfaces et applications musicales qui résultent de nos travaux.

# PUBLICATIONS

Parts of this thesis along with some ideas and figures have appeared previously in the following journal publications :

## INTERNATIONAL JOURNALS

2012 **Esling Philippe**, Agon Carlos "Time series data mining and analysis", *ACM Computing Surveys*, vol. 46, no. 1, 2013.

2012 **Esling Philippe**, Agon Carlos "Multiobjective time series matching for audio classification and retrieval", *IEEE Transactions on Speech Audio and Language Processing* 2013 (Accepted - Major changes).

2012 Hacbarth Benjamin, Schnell Norbert, **Esling Philippe**, Schwarz Diemo "Composing Morphology: Concatenative Synthesis as an Intuitive Medium for Prescribing Sound in Time", *Contemporary Music Review* (to appear)

2011 Lecroq Béatrice, Lejzerowicz Franck, **Esling Philippe**, Baerlocher Loic, Farinelli Laurent, Pawlowski Jan "Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments", *Publication of the National Academy of Science*, vol.108, no.32, pp 13177-13182, August 2011.

## BOOK CHAPTERS

2010 **Esling Philippe**, Carpentier Grégoire, Agon Carlos "Dynamic Musical Orchestration using Genetic Algorithms and a Spectro-Temporal Description of Musical Instruments", Lecture Notes in Computer Science, vol. 6025, EvoApplications Part II, 2010.

## INTERNATIONAL CONFERENCES (WITH REVIEW COMITEE)

2010 **Esling Philippe**, Agon Carlos "Composition of Sound Mixtures with Spectral Maquettes", Proceedings of the International Computer Music Conference, New York, USA, 2010.

2010 **Esling Philippe**, Agon Carlos "Composer les mélanges sonores avec les maquettes spectrales", Actes des 10emes Journées d'Informatique Musicale, pp. 5-15, Rennes, France, 2010

## NATIONAL CONFERENCES

2010 **Esling Philippe**, Agon Carlos "Time series analysis, sound mixtures and orchestration", Presentation in CNRS Japanese-French Laboratory of Informatics, Tokyo University, 2010.

2009 **Esling Philippe**, Agon Carlos "Orchestration and Sound Mixtures", Journées Jeunes Chercheurs en Acoustique Audition et Signal, Marseille, 2009.

*I choose my friends for their good looks,*
*my acquaintances for their good characters,*
*and my enemies for their intellects.*

— Oscar Wilde


# ACKNOWLEDGMENTS

Many thanks to Carlos “Jefe” Agon, Karim “El Guapo” Haddad and all my friends at IRCAM laboratory. It has been three marvelous years that I should not forget ...

# CONTENTS

# Part I

# INTRODUCTION

# 1 THE ARTISTIC PROBLEMATIC

The onset of this study takes its roots in musical orchestration. Orchestration is the subtle art of writing musical pieces for the orchestra, blending the sounds of diverse instruments together by taking into account the acoustic specificities unique to each. The origin of this art lies in the willingness of composers to empower the orchestra to become an expressive unity that can accurately transcribe emotions. This goal can be reached through the knowledge and educated use of the different spectral qualities of each instrument. Composers can then build and adjust particular emotional effects over time. Piston [282] wrote in his treatise that orchestration "*[...] is aimed at discovering how the orchestra is used to translate a musical thought. It is a means to study how the instruments blend together to create equilibrium of sounds*". If we consider the range of expressivity offered by a single instrument, we can get a glimpse on the extent of sonic possibilities offered by an orchestra. Furthermore, given the variety of notes, playing modes and dynamics that can be obtained by an instrument, we can clearly see the combinatorial complexity that is embedded in the art of musical orchestration. Beyond its traditional sense, orchestration rely heavily on the concept of sound mixtures, which ubiquity range nowadays from orchestral to electronic music. Orchestration can be thought as the realm of musical writing in which the timbre acts as the main parameter (we shall try to provide a definition of the timbre in Section 2.3). Amongst all the components of music writing, orchestration has long remained in his teaching as in its practice, an empirical activity. Only quite recently has risen the idea of computer-aided orchestration, which is faced with the vastness of its field of study. Indeed, the topic of orchestration encompasses notions from auditory perception, music analysis, music theory, composition, signal processing and computer science. The difficulty of finding a rigorous formalism along with its youth in the musical discourse still makes orchestration one of the ultimate musical dimensions that have not been studied enough to fit its complexity. We must, therefore, start by investigating the reasons behind this paucity of researches by looking into the past of orchestration in order to envision its future.

## 1.1 FROM THE PAST EMPIRICISM ...

In his earliest essay, Berlioz [39] laid down the foundations of what would later become the teaching of musical orchestration. Already at his time, he noted that the art of orchestration is "*[...] taught as little as the ability to find beautiful songs, beautiful successions of chords and original and powerful rhythmic forms*". Even if we can clearly define a song, a succession of chords or a rhythm, it appears unfeasible to interpret objectively the beautiful nor the original. A few number of treatises followed his work, ongoing to the almost encyclopedic work of Koechlin [206]. From these seminal studies to contemporary works such as Piston [282] and Casella [73], the problem of orchestration is still always exposed and taught to composers on an empirical basis. As there are no fixed, generic rules, such as those that govern harmony in Western music, the exploration of combinatorial possibilities offered by instrumental properties is always reduced to a series of “orchestral recipes”. Thus, even in the most recent works, we can learn that the flute is an instrument with typical pastoral shades or that the bassoon produces tones that resemble a libidinous old man [73]. As strict, universal orchestration rules continuously seem to have eluded scholars over several generations, the authors of treatises delve in producing a collection of examples drawn from their own experience and explore the resulting orchestral color in their own terms. The various treaties thus boil down to a series of recipes identifying some of the orchestral archetypes. It is clear that these empirical approaches of orchestration preclude the systemic use of such treatises on a scientific basis.

In his practice of orchestration, the composer must face the choice of using his own (necessarily limited) personal experience or to turn to orchestration treatises that fail to examine the combinatorial possibilities of orchestral timbre on a rigorous basis. As full of examples as might be all the treatises, the question of their accuracy and scope, however, deserves to be asked. First, as exposed before, treatises lack the exploratory component that could provide a structural support on instrumental mixtures for the composers to adjust the final outcomes. Furthermore, recent developments in instrumental music have focused on the introduction of exotic playing modes that allow instruments to produce sounds previously unheard. Therefore, as most of the orchestration studies date back to several decades, there seems to be a potential obsolescence of their repertoire of the study. The language used by various authors may also already reflect the aesthetic vision of their time. From all these observations, it would seem that trying to devise an orchestration treatise on rational grounds is bound to be a perilous operation.

## 1.2 ... THROUGH MUSICAL WRITING ...

Even if the art of orchestration focus on the use and evolution of sound properties, it remains an act of musical writing. In fact, it is extremely difficult to separate these two aspects as a musical material is often thought and written for a pre-defined set of instruments. However, orchestration cannot be enclosed in a single frozen time of writing, or limited to the field of instrumental technicalities. Apart from isolated chord situations where the pitches of all instruments are stable, every other musical context (figures, textures, gestures) seem to fall together within a tied writing and orchestration. In its daily practice of orchestration, the composer does not only superimpose stationary sounds but instead focus on carving, vertically and horizontally the sound material. He can simulate the attack of an instrument by another, with a third one acting as resonance. The position of the orchestration in relation to the entire musical knowledge have significantly evolved through different eras and authors. Koechlin, for example, seems to consider it as a technique totally subservient to the other dimensions of writing: "*We must carefully plan every orchestral element such as accents, cadences, progressions and dynamic overlaps in relation to a piece as a whole. Each change should be based on its musical context*" [206]. From these observations, we can distinguish two modes of writing for the orchestra. First, a "holistic" orchestral writing where melodies and instruments are thought of as indivisible units, which we call *inductive orchestration*. On the other hand, an "abstract" writing in which the score and chords are produced independently and then orchestrated, which we call *projective orchestration*. Koechlin advised against this practice, stating "*the too much prevalent belief that the orchestration is the basic allotment of timbres in various instrumental lines is clearly inadequate*" [206].

- *Inductive orchestration* is created by having precise orchestral colors and effects in mind and trying to produce combinations of instruments that could achieve these ideas.
- *Projective orchestration* is produced by first writing an "abstract" score and then determining the allocation of different melodic lines to instruments. Although Koechlin decried this practice, famous examples include Maurice Ravel who orchestrated his piano plays and, therefore, had not thought about the instrumental colors when writing his original pieces.

From these two potential forms of orchestration, composers seem to be more keen on inductive orchestration, thus writing musical pieces with a precise idea of their orchestral colors. In both cases, the art of orchestration is directly related and even almost indivisible from musical writing, as they are not isolated processes from the compositional practice. Once again, the problem of music writing is an intricate area for scientific research that encompass techniques that fall under subjective assessments, and whose mechanisms appear difficult to formalize.

## 1.3 ... TOWARDS A MODERN TREATISE ?

Given this empirical tradition, it seems almost hazardous to tackle the orchestration on a rigorous scientific basis. However, Bregman [57] envisioned, "*[...] it should be possible to write an orchestration treatise based on fairly abstract principles to be applied to all styles of music. This would not comprise a collection of precepts accumulating imperatives and prohibitions, but a guide to how to get a particular sound, leaving the composer free to alter the outcomes*". When Bregman refers to *all* styles of music, we can see that the breadth of musical orchestration goes beyond instrumental writing and can be generalized to the notion of sound mixtures, mainly when considering the latest trends in contemporary music.

The advent of electroacoustic music has produced a fundamental shift in instrumental music, which seems now to evade from the traditional categories of musical writing (*pitch, duration, mode*) and heads itself towards the composition of inharmonic and noisy sounds which exhibit large timbre variations over time. The most evident marks of this evolution are the *spectral music* (Grisey, Murail), *concrete instrumental music* (Lachenmann, Sciarrino) and the *school of complexity* (Ferneyhough, Dillon) which seek to "*saturate writing*". It is interesting to note that this problem seems specific to Western music, locked in his writing since the Gregorian era. Indeed, the music from other cultures, mostly passed down through oral traditions has long been attached to address these categories of complex and noisy musical elements. Amongst nowadays composers, many are focusing more resolutely towards the expressive qualities of sound and the potential of complex sounds and noises offered by the infinite capabilities of electronic source materials. Musical orchestration in the context of these sonic possibilities becomes an even more complex operation. This question brings us back to the fuzzy boundaries that exist between sound and noise. With the ceaseless flourishing of electronic music, it seems that this limit has been pushed further into a corner. The advent of technology has made us consider noisy sounds as potential units of musical creativity, evidenced by the approach of Pierre Schaeffer [317]. He chose to compose with sounds and noises recorded from any potential source in the every day environment, thus giving rise to the *musique concrète* movement. He exposed the need to "*[...] replace the limited variety of instruments that constitute an orchestra with the* *infinite variety of timbres provided by noises obtained through special mechanisms*". Schaeffer suggested elevating noises to music by transforming and repeating such sounds in order that the listeners' awareness might be directed by temporal expectations. This approach thus seeks to create a causal dependence between elements in order to delineate a clear boundary between music and noise. These methodologies show us that the process of organizing sound properties can be highly conditioned by the elementary units that are sought to be combined.

## 1.4 ELEMENTARY UNITS OF ORCHESTRAL STUDY

As discussed previously, the art of musical orchestration embeds all the techniques that attempt to combine musical elements and "*[...] consider how they can contribute to an emotional effect in their musical association*" [39]. In the classical orchestra, we can perform a categorization of these elements on several scales, starting from the instruments (violin, cello, clarinet, flutes), the families of instruments (strings, woodwind, brass) and up to the mode of production. However, even inside a common production apparatus, the sound characteristics of each instrument provide an almost infinite variety. A tremendous scope of expressivity can be found in each instrumental repertoire of playing modes, range of notes, intervals and chords, range of dynamics and finally the overall structure and rythm that creates a musical context. Even with all these parameters fixed, the nature of instrumental sounds still remains somehow stochastic. Thus, we must keep in mind that even at the same pitch, dynamics and duration, the resulting signal can vary from one instance to another, which further deepen the combinatorial complexity of the orchestral study. This variability of elements appears through different dimensions and can be observed on multiple temporal scales. Many of those are still to this day oblivious to scientific study. Nevertheless, in order to combine these elements appropriately, we need to consider their range, playing modes, playability, dynamics, articulations, speed, role and purpose in an ensemble and even their individual ability to translate musical thoughts. As underlined by Piston [282], "*[...] we can not overstate the importance of instrumental knowledge. An appropriate instrumental writing is undoubtedly the determining factor in the success of an orchestration*". Thus, it seems essential to take steps towards the support of the modern instrumentarium in which the computer seems to have an evergrowing importance. If the musical research does not capture these individual sound particularities as an aim in itself, the orchestration will remain an empiric and haphazard science. As we try to devise a formalism to pave the way for modern orchestration, we thus need to go through with a deep understanding of the sound properties of its elementary units.

## 2 STUDYING ELEMENTS

In the study of musical orchestration, a fundamental question lies in the selection of elements to combine. Indeed, how to integrate the properties of different sounds, if we are not yet able to master these constituents and how they are perceived.

### 2.1 MUSICAL ATOMS

The study of orchestral pieces can be performed on several distinct scales as exemplified in Figure 1. The infinite variety of instrumental combinations can generate widely varying sound results, therefore, inducing an aspect of *combinatorial complexity*. At the same time, the temporal evolution of musical structures strongly shape the perception of musical pieces. These *macro-temporal structures* can create *tension*, *expectation*, *release* and a wide variety of emotions depending on their construction. However, if we focus our attention to even a *single* element inside these complex structures (such as the red-circled note in Figure 1), another wealth of complexity reveals itself. These elements, that we will call *musical atoms*, also exhibit unique temporal evolutions of their spectral properties. These *micro-temporal properties* vary even for the same instrument at the same pitch and loudness, as we shall discuss in Section 2.3. Furthermore, the temporal evolution of multiple decorrelated sound properties can be perceived simultaneously leading to a *multidimensional perception*, which we discuss in details in Section 3.3.

We call these elements *musical atoms* in the broadest sense that they constitute a coherent spectral unit. The first element that should come to mind is a single note from an instrument, but a continuous glissando also form a logical unit. Therefore, we also embed in this definition every acoustic entity that represent a *unit of musical creativity*. Indeed, as exemplified by the approach of Pierre Schaeffer [317], even everyday noises can be elevated to the rank of musical units depending on the temporal organization of such elements.

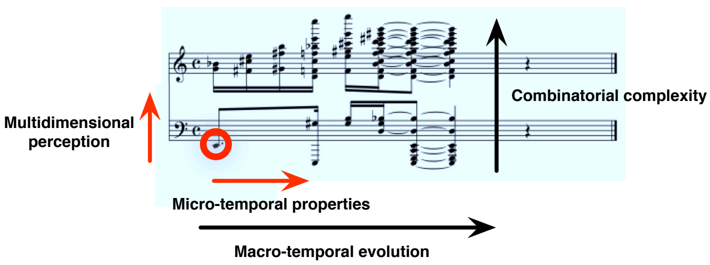



Figure 1: Different levels of complexity appear in the study of music, where the processing of musical pieces is performed on several scales simultaneously. The various potential combinations of instruments induce a *combinatorial complexity*. The *macro-temporal evolution* of these musical structures strongly condition our perception. However, if we focus on a single musical atom, it also embeds *micro-temporal* evolutions of its spectral properties that are perceived in a *multidimensional* fashion.

## 2.2 SIGNAL AND SYMBOLISM

Composition may be seen as the act of projecting a hypothetic signal that exists only in a composer's mind to an efficient symbolic representation. This strategy has been used for centuries as it allows sharing this mental image to performers that will try to reproduce it in the most accurate way. Western music was largely based on this *harmonic paradigm*, i.e. it built its musical concepts from sounds considered harmonic and stationary. Western musical notation is also without coincidence, consistent with this paradigm (writing a pitch as a point is reducing a sound to its fundamental, and considering that, inside this sound, changes in the harmonics follow a parallel and constant evolution). However, this symbolic tradition has now to coexist with a consideration of composers increasingly marked with the spectral qualities of musical elements. The musical evolutions have led us to reconsider the inharmonic and noisy sounds as part of the musical perception. A substantial theoretical debate in the study of sound mixture composition comes from the use of multiscale representations for linking heterogeneous data (cf. Figure 1). From this, unfold a contemporary issue in computer music research : the *signal / symbolic interaction*. These two research streams have long remained impervious to each other, partly because of the apparent heterogeneity of their objects of study. On one hand, the analysis and synthesis of digital signals contributed to the comprehension and production of sounds previously unheard. On the other hand, the symbolic approach focused on the analysis of musical notation structures, but the composition of sound mixtures is located precisely at the intersection of these two lines of research. If it claims to produce timbres, it is through a symbolic writing process. Therefore, it is an ideal meeting point between the symbolic and spectral domains. Sound as a material coexists and interacts with formal structures. Nowadays, computers offer the possibility to manipulate a sound object in a compositional process, while simultaneously studying its acoustic properties through analytical tools. We can thus combine the symbolic discipline of writing with the spectral domain of timbres possibilities. It should, therefore, become possible for the composer to relate the exploration of sound to the organization of symbolic data.

## 2.3 MULTIPLICITIES OF TIMBRE

As stated earlier, the *timbre* is the dominant parameter of this study, as musical orchestration is focused towards attaining new timbre through the combination of individual instrumental timbres. However, we carefully avoided defining and overusing this term, and preferably used the instrumental *sound properties* or *qualities*. Indeed, when the musical timbre is put forward in a scientific study, the first problem to arise is that of its precise meaning. It seems that a consensual definition of musical timbre has eluded scholars for several generations. Attempting to formulate a unique and accurate definition inevitably leads to a number of difficulties, as the systematic use of this word has finally made it airtight and elusive. Despite a century of consecutive studies, there seems to be no consensus on a comprehensive definition from which researchers could build models or theoretical methods. The timbre rather encompass a multitude of definitions that are formulated differently by acousticians, musicians or computer scientists, as we will detail in the remainder of this section.

The American Standards Association (ASA) defines timbre as "*[...] that attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar*" [16]. From this definition, the timbre would seem like a property that could intervene only in discrimination tasks but can not be quantified positively. In simpler terms, we could say that the timbre of an instrument is its collection of sound *characteristics* which allows to recognize it and makes it differentiable from other instruments even at the same pitch and loudness.

### 2.3.1 *Timbre and acousticians*

Following the work of Joseph Fourier on the decomposition of periodic functions in 1822, the first significant research on the instrumental timbre can be traced back to the seminal work of the physicist Hermann Helmoltz [167] in 1863. At this time limited by his means of investigation, he is forced to restrict his analysis to sustained sounds, which he called *musical tone*. By using a series of resonators, he evidenced the harmonics of instrumental sounds and then suggested that the timbre is defined by the mean intensity of the harmonic components. However, he already noted "*[...] with a little thought we also see now that some of these acoustic features depends on how the sounds start and finish*". Following the work of Helmholtz, the study of instrumental sounds has long remained entirely ignorant of their temporal aspects. This approach is now strongly avoided (as we will discuss further in Section 2.4) as it is widely recognized that the perception of timbre can not be separated from its temporal aspects. Following the work of Risset [306], it is now well-accepted that the characteristics of timbre are heavily influenced by the temporal variations of *each* harmonic component. His study also revealed the dynamic nature of the energy envelope and the importance of the attack segment in recognition of sounds. Therefore, we can define the *causal timbre* that is involved in recognition of the sound source. On the other side lies the sound properties that provide a qualitative perception of sound. The timbre is thus both the identity of the sound source and sound qualities it possesses.

### 2.3.2 *Timbre and musicians*

It seems that the appearance of the timbre in the musical discourse comes from the work of Berlioz [39]. He introduced his treatise by saying : "*The purpose of this book is first, an indication of the extent of the instrumental mechanisms. Then we shall turn to the so far neglected nature of the timbre, the nature and expressive abilities of each instrument and the best known methods for grouping them properly*". Thus, when defining timbre, the composers take into account the character of each instrument along with its expressive capabilities. In the history of Western music, the instruments that were originally designed as substitutes for the human voice, have progressively became "*voices*" on their own, whose identity favored polyphonic sound listening. The instruments then started to be employed as entities providing variations of colors in musical discourse. In this musical context, it appears that the two complementary aspects of the timbre are still an integral part of musical creation: the *identity* of a recognizable part and the *evolutions* of its colors. Furthermore, when considered in terms of music writing, timbre plays a dual role at the two poles of the instrumental universe which are the *articulation* and *fusion* [57]. The *articulation* of music is created by the temporal evolution of its structure, based on the possible identification of individual timbres. On the other hand, the *fusion* is the effect that can make it impossible to determine the individual components of a sound mixture. It is thus intended to lead listeners towards new elements of the musical discourse.

Until the 19th century, the timbre seems to have played an almost decorative role in musical composition. Romantic composers such as Hector Berlioz began to consider the timbre as part of the expressive power of music. Arnold Schoenberg and the Vienna School tried to expand its influence in the music discourse. Contemporary composers now use the timbre as a central element of musical aesthetics. Unfortunately, despite their interest in the timbre, composers might not take advantage of its ubiquity, as the timbre is created through complicated physical phenomena which are, therefore, difficult to describe and use.

### 2.3.3 *Timbre and computer scientists*

Through all the potential definitions of timbre, the notions of sound *qualities* or sound *characteristics* are always prevalent. Therefore, over the past decade, a tremendous amount of work has been devoted to finding relevant high-level features to compute over sound signals. These researches have been made possible by advances in computation speed of the Fourier transform through the Fast Fourier Transform (FFT) algorithm. Over the years, an increasing number of audio features have been developed in order to provide a characterization of the timbre by describing it through a set of complementary features. These can be coarsely divided into six main categories; ie. *energy*, *frequency*, *harmonic* (computed only on the harmonic peaks), *spectral* (computed on the whole distribution of the spectrum), *noise* (computed after removing the harmonic peaks) and *perceptual* (computed after filtering the signal with a model of the human ear). Each of these features represents the temporal evolution of a particular characteristic of the related sound. It has now become straightforward to compute this whole set [274] over a sound signal in order to obtain different aspects of the sound. Psychoacoustic studies have further shown the correlation between these sound features and perceptual dimensions. One of the best-known example is the *spectral centroid* that has been shown to be related to the *brightness* of sounds [179]. It has also been shown that the *attack time* heavily influence the perception of percussive aspects of a sound [144]. Other perceptual dimensions have been studied like *sharpness* that relates to the position of *harmonic energy* in the spectrum [369] or *tension* that seems to be primarily connected to *roughness* [288]. Hence, in the computer scientist approach, the timbre is sought to be fully characterized by a set of audio features as comprehensive as possible.

## 2.4 TIME SCALES CONTINUUM

From a purely aesthetic perspective, it appears obvious that a collection of beautiful things placed upon another with no structure is clearly inadequate. Thus, music can not be represented by a collection of "beautiful" instants frozen in time but is rather developed through a carefully planned temporal structure and progression. We have seen, through the approach of Schaeffer, that a collection of unpleasant elements can be transformed into music if they are given the proper temporal logic. Every composer knows that the sound of a note depends on its context and that it is the sound pattern, not the isolated note, which determines the auditory perception.

The previously stated definition of timbre by the ASA in 1960 [16] is accompanied by a note (p.45) which states, "*[...] timbre depends primarily on the spectrum of the stimulus, but it also depends on the waveform, sound pressure, spatial position and temporal characteristics of this stimulus*". It is now well-accepted that the timbre does not only depend on the average spectral distribution of sound spectra, but is also strongly tied to their temporal characteristics. Risset [306] found that the proportion of harmonics in the spectrum of brass instruments increases with their loudness. This shows that the timbre can not be described by the static values of many parameters, but is instead characterized by the temporal evolution of some of these parameters. This is also confirmed by several works on the perception of timbre [56, 256, 319] that demonstrate the importance and primacy of temporal descriptions. As argued previously, the spectral characteristics of timbre are dynamic and continuously evolving elements.

From all these observations, it appears that time should be the primary topic of research in any musical approach, even at the smallest temporal scales. If orchestration requires tools for the vertical and horizontal arrangement of sounds, we have to formalize processes that evolve over time. The revolution of the twentieth century by the emergence of the timbre and discovery of other cultures urge to overpass in one way or another, the paradox of writing for harmonic instruments. In order to accompany this aesthetic paradigm shift, it seems essential to better understand, describe, analyze and compare complex sounds by firstly taking into account the temporal aspects of timbre.

# 3 PERCEIVING ELEMENTS

In order to study the potential approaches to combine musical atoms we have seen that we must first understand what we can and can not perceive, but most of all *how* we perceive them. All along this discussion, we will take an analogy with the visual perception in order to clarify our ideas to the readers.

## 3.1 THE DOORS OF PERCEPTION

The physical perception of living beings has been shaped through millions of years of genetic evolution. This physiological shaping is based on a response to survival needs and for the beings to be able to understand, identify and interact with their environments. The continuous and consistent modifications of the sensory systems can be seen as an indirect action, in which the living species are confronted to various environments to which they should adapt to survive. Hence, what we perceive is based on the evolution of our needs in our environment. Therefore, all forms of perception are meant to provide an *organization* and possible *interpretation* of the surroundings. If we look at the limits of the visual perception, presented in Figure 2, it seems that our visual system is limited to an extremely narrow segment of the electromagnetic spectrum. However, in this small fraction of sensory information that we call *visible light*, we can still distinguish millions of colors and perceive extremely complex and detailed spatial structures.

We now turn our attention to the limits of our auditory perception, presented in Figure 3. As for visual perception, we can only perceive a very small fraction of the acoustic spectrum. However, in this narrow portion of the air vibrations, we are able to differentiate hundreds of pitches and perceive complex temporal evolutions and structures. Finally, even if we consider a fixed part of this acoustic spectrum (supposedly here instruments at the same fixed pitch), we are still able to differentiate between all these sources based on their various properties. Hence, to take a computer scientist's analogy, we could say that even if our hardware is somehow limited, our software seems pretty efficient for making the most of this restricted information. Therefore, our sensory systems are meant to provide an *organization* of our surroundings, but it is the *interpretation* of this sensory information that enable our perceptual systems to achieve *identification* of their constituents. Therefore, perceiving elements is being able to identify elements. In turn, in order to be able to perform these distinctions between elements, we need to rely on a complex assessment of their (dis)similarities.

## 3.2 COMPLEX AND MULTIFACETED SIMILARITIES

### 3.2.1 *Assessing visual similarity*

We start by taking a visual example in order to make explicit the difficulties in defining the notion of similarity, that we will subsequently apply to sound perception. The analysis of visual similarity is presented in Figure 4. If we consider a set of three basic visual units (*square, circle* and *triangle*) and use only the three elementary colors (*red*,

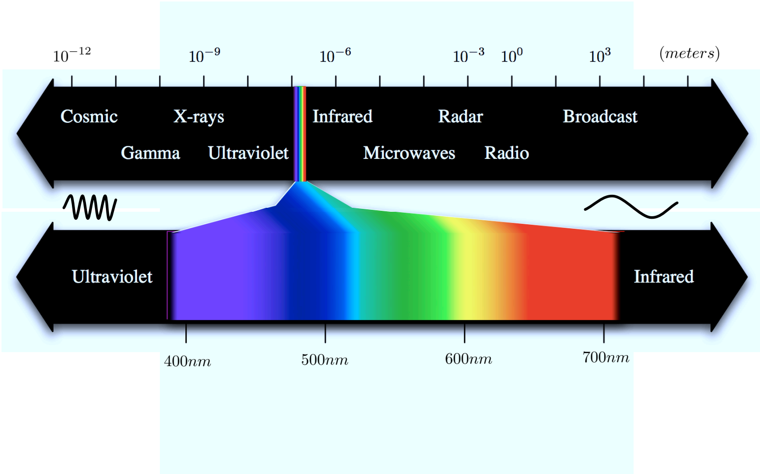



Figure 2: The doors of our visual perception. We are able to perceive only a very small fraction of the electromagnetic spectrum. However, even in this narrow perceivable part we can still differentiate between millions of colors and perceive extremely complex and detailed spatial structures.

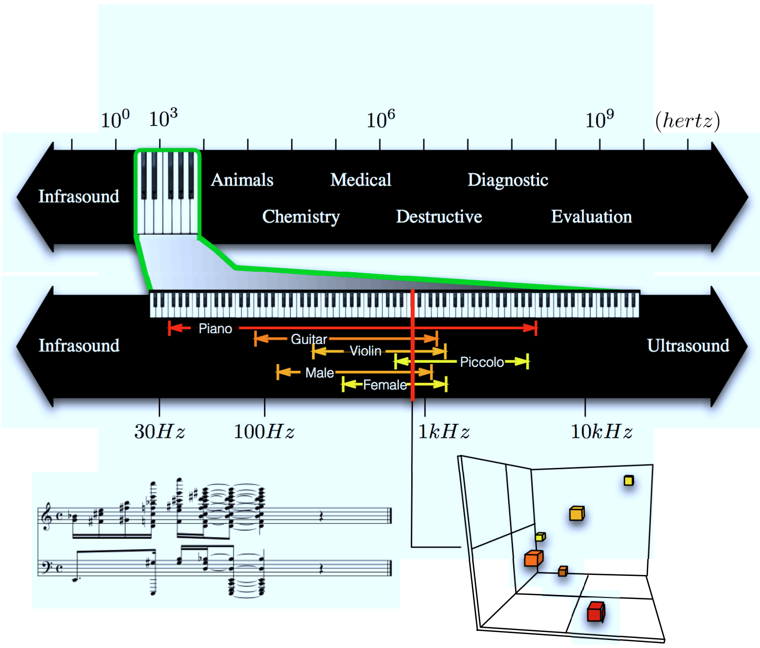



Figure 3: The doors of our auditory perception. As for visual perception, we can only perceive a very small fraction of the acoustic spectrum. However, in this narrow portion of the air vibrations, we are able to differentiate hundreds of pitches and perceive complex temporal evolutions and structures. Finally, even if we take a fixed portion of this acoustic spectrum (supposedly instruments at the same fixed pitch), we are still able to differentiate between all these sources based on their various properties.

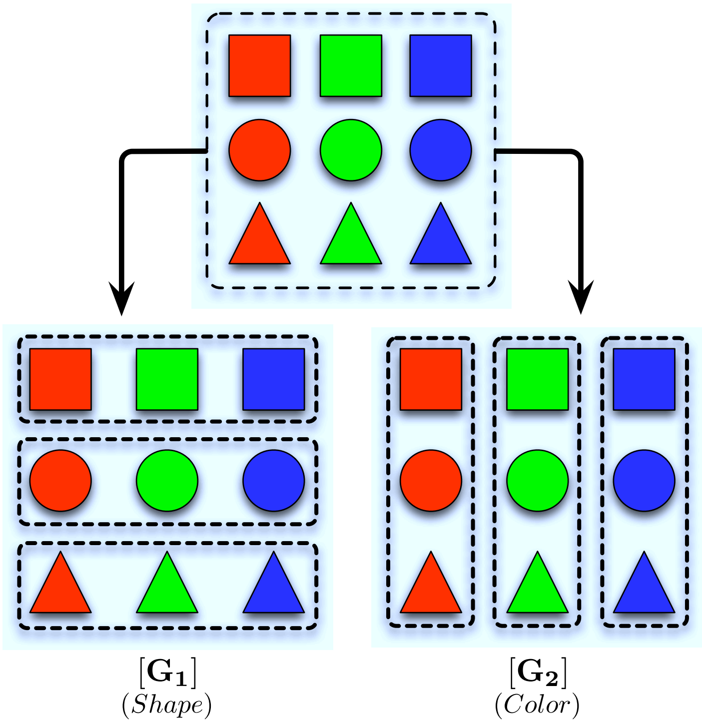



Figure 4: The assessment of *visual* similarity in a set of elementary units. Given this set of elements, we are faced between evaluating their similarity based on the *shapes* of elements such as grouping $\mathbf{G_1}$ or rather based on their *colors* such as grouping $\mathbf{G_2}$.

*green* and *blue*), we can obtain a set of nine elementary units. Given this set of elements, we are faced between two forms of similarity assessments. We can either process the similarity based on the *shapes* of elements such as grouping $\mathbf{G_1}$ or rather based on their *colors* such as grouping $\mathbf{G_2}$. Being aware of these two features when trying to process the similarity makes it almost impossible to provide a *single* measure of similarity between these elements. For example, if we try to rank the similarity between the red circle and all other elements, it seems unfeasible to determine if the most similar item is one of the circles or another shape that is filled with the same color. This single similarity score (ie. finding *the* most similar element) can not be provided without putting some *preferences* over the two dimensions of variability (*shape* or *color*).

However, this case already allows to get a first sight on what could be the notions of *subjective* and *context-dependent* similarity ratings. Indeed, if this document has been printed in black-and white (and same goes for color-blind people reading it), then the $\mathbf{G_2}$ grouping becomes less relevant, and the similarity can be processed solely through the shape of objects. Therefore, the similarity between elements might be assessed following different *directions* that can vary depending on the subject and the context.

### 3.2.2 *Assessing sound similarity*

Now let us turn our attention to the assessment of sound similarity. In order to do so, we consider the most elementary sound that could ever be synthesized or heard, namely a *sinusoidal signal*. This primary unit that forms the basis of all harmonic sounds can be synthesized by computing the temporal signal

$$x(t) = \mathcal{A}(t) . \sin\left(2\pi\mathcal{F}_0(t) . t + \phi_0\right) \tag{3.1}$$

with $\mathcal{A}(t)$ the *amplitude* function over time (the "*loudness*") and $\mathcal{F}_0(t)$ the *frequency* function of the sinusoid (its "*pitch*"). In our everyday life, we can easily differentiate the pitch and loudness of sounds, so it is fairly painless imagining a sound with an increasing pitch and decreasing loudness simultaneously (for instance imagine a squeaking door or an ascending whistle moving away from you). So we are here studying the simplest element that we could ever find in audio processing, through its most basic properties. A graphical interpretation of this assessment is presented in Figure 5. We can easily define two different temporal functions for the *loudness* and *pitch* properties. We suppose that the temporal evolution of each feature can either be *descending* or *ascending*. Therefore, the temporal evolution of the amplitude $\mathcal{A}(t)$ can either be $\mathcal{A}^1$ or $\mathcal{A}^2$ and the temporal evolution of the frequency $\mathcal{F}_0(t)$ can either be $\mathcal{F}_0^1$ or $\mathcal{F}_0^2$. Given these two possibilities for each feature, we can synthesize the set $\mathcal{S}_{\mathcal{A}}^{\mathcal{F}_0}$ of four different sounds $\mathcal{S}_1^1$, $\mathcal{S}_2^1$, $\mathcal{S}_1^2$ and $\mathcal{S}_2^2$. Now if we try to determine the similarity between elements inside this set, we face the same problematic as for visual similarity. We can either evaluate the similarity between elements based on the evolution of their *loudness* properties such as grouping $\mathbf{G_1}$ or rather based on the temporal evolution of their *pitch* such as grouping $\mathbf{G_2}$. As discussed in the previous section, it seems impossible to provide a *single* measure of similarity without imposing some *preferences* over the two dimensions of variability.

## 3.3 multidimensionality of timbre perception

As we discussed earlier, several psychoacoustic studies have shown the correlation between sound features and perceptual dimensions. The previous example also showed us that we are able to perceive the temporal evolution of decorrelated audio features simultaneously. This multidimensionality of timbre perception has been extensively demonstrated in psychoacoustic studies [148] through the concept of multidimensional timbre spaces. Several authors had already pointed out that timbre is a multidimensional phenomenon [283, 250], and that our perception organizes these dimensions given a sound context [376]. Most of the psychoacoustic studies that can be found in the literature often makes use of these multidimensional spaces in order to distinguish different sounds [316, 376]. Timbre has even been said to *"[...] tend to be the psychoacoustician's multidimensional wastebasket category for everything that cannot be labeled pitch or loudness"* [244]. Authors in the Music Information Retrieval (MIR) community have also pointed out the multifaceted nature of audio perception [110] and that a single measure is unlikely to convey the perceptual similarity of audio signals [363]. Sound retrieval systems should thus be flexible enough so that depending on listeners and target timbres, variable importance could be put on different sound properties during perceptual similarity evaluations [245], but yet no current audio-retrieval system seems to address these limitations.

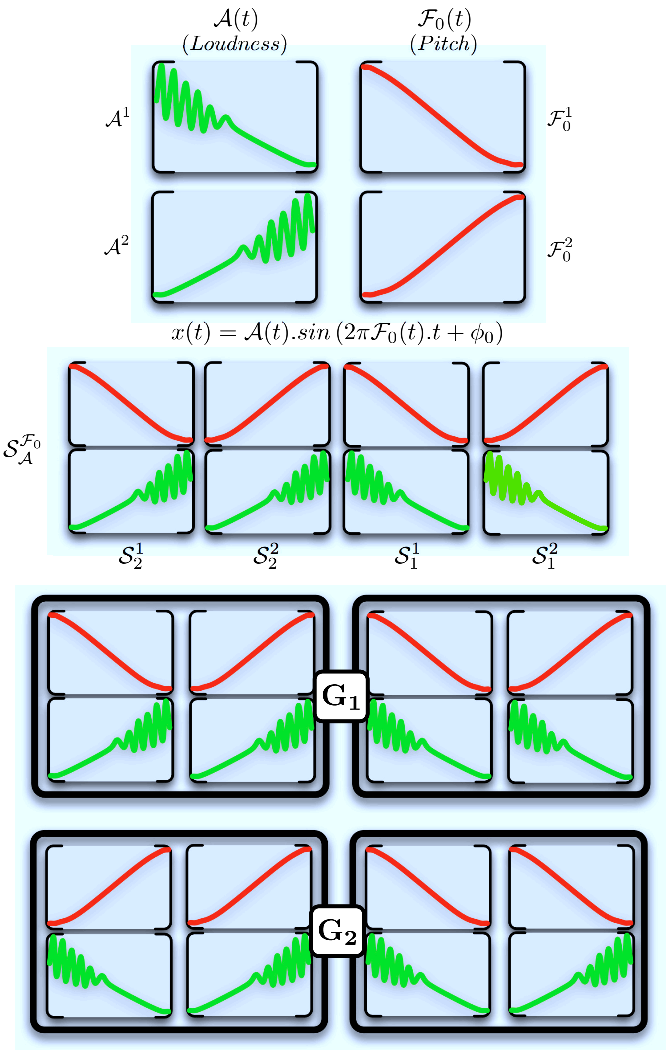



Figure 5: The assessment of sound similarity given its most elementary unit, a synthesized sinusoidal signal. For this signal, we can define the temporal function of its *amplitude* ("loudness") and its *frequency* ("pitch"). Therefore, we define two very simple temporal functions for each feature. The loudness $\mathcal{A}(t)$ can either be set to $\mathcal{A}^1$ or $\mathcal{A}^2$ and the pitch $\mathcal{F}_0(t)$ can either be set to $\mathcal{F}_0^1$ or $\mathcal{F}_0^2$. Given these two possibilities for each feature, we can easily synthesize the set $\mathcal{S}_{\mathcal{A}}^{\mathcal{F}_0}$ of four different sounds $\mathcal{S}_1^1$, $\mathcal{S}_2^1$, $\mathcal{S}_1^2$ and $\mathcal{S}_2^2$. If we try to assess the similarity between elements inside this set, it seems unfeasible to choose if we should to group elements based on the evolution of their *loudness* properties such as grouping $\mathbf{G_1}$ or rather based on the temporal evolution of their *pitch* such as grouping $\mathbf{G_2}$.

# 4 PUTTING THE PIECES TOGETHER

## 4.1 RATIONALE OF THIS STUDY

The previous chapters discussed the artistic problems that pose the questions of this study. Musical orchestration is the art of combining the timbre of instruments in order to achieve a musical thought. While discussing the intrinsic properties of what we called *musical atoms* to be combined, we saw that we should especially take care of two concerns. First, the temporal evolution of audio properties should be a central element of this study. Second, we discussed the extent of our auditory perception which clearly shows that we are able to discern several decorrelated temporal properties simultaneously. Based on these observations, we anticipate a compelling question that is yet to be answered. In order to obtain a true assessment of the timbral similarity we need to take into account both the multidimensional aspect of timbre and the temporal evolution of its structure. However, to the best of our knowledge, it seems that such an approach has never been undertaken. Therefore, this study focuses on providing a flexible framework that can assess the *temporal* similarity between elements on several dimensions, without merging the similarities into a single distance measure.

We strive to give here a brief discussion on the epistemological considerations underlying this study. Trying to tackle a complex artistic problem raised a series of perception issues. From these questions, appeared the need for a paradigm yet to be solved. We will, therefore, focus on trying to formalize this unsolved problem and will present algorithms to solve it. We will see that avoiding to merge the distance measures between the temporal features of each dimension can provide a compelling framework of study. Thus, we will show that the applications of this model can go far beyond the realm of music and audio processing. By applying these concepts to generic classification problems, we will show that these auditory ideas can provide an improvement on a variety of scientific fields. The path that we follow in this study, show us that the artistic problems can raise questions on a wider scale. Trying to answer these questions can in turn provide comprehensive and powerful approaches that can be beneficial to a variety of scientific fields. Hence, music can help us to study science through the complexity of the problems it contains. We will show that we can also close this epistemological loop, by using all the knowledge gained from these studies in order to solve the first problematic. Figure 6 present this conceptual loop of study.

## 4.2 SCIENTIFIC CONTRIBUTIONS

We explore in this thesis a wide variety of scientific topics, therefore, we try here to summarize our contributions along this study.

- While reviewing the field of time series data mining, we propose four axioms of robustness through which the robustness of any time series distance measure can be evaluated.

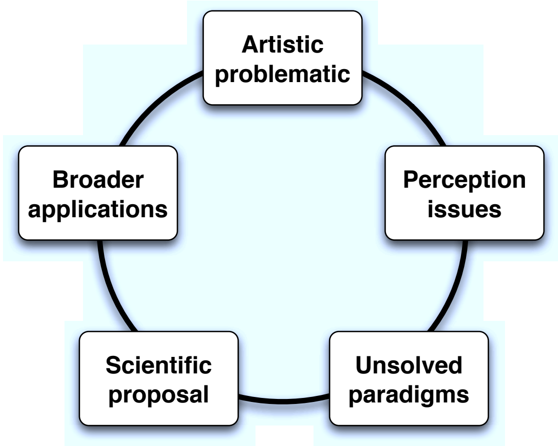



Figure 6: The *epistemological loop* of this study.

- We introduce and formalize the novel problem of *MultiObjective Time Series* (MOTS) matching which allows to take into account both the multidimensional aspects of elements and their temporal structures. We propose two efficient algorithms to solve this problem with a sublinear time complexity. We validate this framework in the field of audio perception through perceptual studies.
- Based on the MOTS framework, we introduce two innovative audio querying paradigm that provide more intuitive interactions with sound samples. We further validate these paradigms with an extensive usability evaluation study.
- We show how to adapt this notion of flexible similarity matching to classification problems. We introduce a new classification selection criteria based on the hypervolume dominated by each class and study this new classification paradigm over a wide range of datasets. By not merging distances into a single similarity measure, we show that our classification paradigm outperforms state-of-art methods on several scientific fields.
- Based on this paradigm and by still drawing inspiration from the art of listening, we show how to construct a system for biometric identification based on the sounds produced by heart beats. In order to do so, we propose a novel set of features based on the Stockwell transform, called *S-Features*.
- We also show some specific audio applications of our classification paradigm where it allows to outperform state-of-art methods on reference audio classification problems.
- We construct a novel computer-aided orchestration algorithm that can take into account the temporal evolution of audio features. We introduce a multi-level segmentation method that can take into account the different segments that are contained *within* an element at smaller time scales. We show that altogether these new orchestration procedures strongly outperforms the previous approaches.

## 4.3 STRUCTURE OF THIS DOCUMENT

ELEMENTS OF TIME We start by giving an in-depth review of the time series data mining field (Chapter 5), through its most representative tasks (Section 5.3). We further divide the relevant literature based on the three main aspects of time series handling, namely *representation* methods (Section 5.4.2), *distance* measures (Section 5.4.3) and *indexing* structure (Section 5.4.4). We then provide a overview of the multiobjective optimization field (Chapter 6) by introducing its core notions (Section 6.1), providing a summary of algorithms classification (Section 6.2) and finally presenting some of its application. (Section 6.3).

MULTIOBJECTIVE TIME SERIES (MOTS) MATCHING We then introduce the problem of *MultiObjective Time Series* (MOTS) matching and its formalization (Chapter 7); we further underline the core differences between this novel problem and multivariate matching (Section 7.2) and also briefly discuss its computational complexity. We introduce two efficient algorithms to solve this problem (Section 7.3) and analyze their relative merits on synthetic and real datasets. We describe the application of this framework for innovative audio querying (Section 7.5) by introducing two new querying problematics in the field of audio samples retrieval, namely the *MultiObjective Spectral Evolution Query* (MOSEQ) (Section 7.6.2) and *Query by Vocal Imitation* (QVI) (Section 7.6.3). We discuss the evaluation of the MOTS framework through perceptual studies (Chapter 7.7) and usability evaluation of the MOSEQ and QVI querying paradigms (Section 7.6.3).

HYPERVOLUME CLASSIFICATION (HV-MOTS) Based on these results, we extend the range of this study and show how to apply these notions of variable similarity evaluation to classification problems by introducing the HyperVolume-MOTS (HV-MOTS) classification scheme (Chapter 8). We show that even within the multiobjective framework that avoids merging distances into a single measure, we can still rank classes by relying on the hypervolume dominated by each (Section 8.1). We discuss the relationship between this novel classification scheme and other distance-based classifiers (Section 8.2) but also to more generic classification schemes (Section 8.3) and discuss its main advantages and drawbacks. We provide a large scale study of the performances of this classification technique (Chapter 9) on a wide range of datasets that covers several scientific fields. We show the statistical superiority of the HV-MOTS classifier over well-established classification schemes (Section 9.3) and state-of-art results on the same datasets (Section 9.4). Based on the HV-MOTS classifier, we show how to build a biometric identification system for heart beat sounds (Chapter 10). We construct it by considering listening as an art (Section 10.3) and developing a specific set of features based on the Stockwell transform, called *S-Features* (Section 10.3.3). We show that using heart sounds as a biometric feature provide a reliable identification (Section 10.4.3) and that this feature is not affected by the phenomenon of *template ageing* (Section 10.4.3) over a time span of two years, supported by the recordings collected in the Mars 500 isolation study. We illustrate the application of the HV-MOTS framework to audio problems (Chapter 13) through generic audio samples classification (Section 11.1) and sound morphology (Section 11.2).

GOING BACK TO MUSIC We show how this knowledge gained through broader applications can be put to use in the field of musical orchestration (Chapter 12). We propose a new orchestration system based on an algorithm that use the MOTS framework and that rely on an entropic segmentation method (Section 12.4). We show that this new algorithm can improve the previous approaches for computer-aided orchestration (Section 12.4.2). We then present other artistic applications of the MOTS framework (Chapter 13).

Finally, we offer our lines of future work (Chapter 14) and conclusions (Chapter 15).

Part II

# ELEMENTS OF TIME AND PERCEPTION

# 5 TIME SERIES DATA MINING

In almost every scientific field, measurements are performed over time. These observations lead to a collection of organized data called *time series*. The purpose of time series data mining is to try to extract all meaningful knowledge from the *shape* of data. Even if humans have a natural capacity to perform these tasks, it remains a complex problem for computers. In this chapter, we intend to provide a survey of the techniques applied for time series data mining. The first part is devoted to an overview of the tasks that have captured most of the interest of researchers. Considering that in most cases, time series tasks rely on the same components for implementation, we divide the literature depending on these common aspects, namely *representation* techniques, *distance* measures and *indexing* methods. The study of the relevant literature has been categorized for each individual aspects. We also introduce in this chapter four types of robustness that we formalize and thanks to which any kind of distance measure could then be classified. Finally, we submit various research trends and avenues that can be explored in the near future.

## 5.1 INTRODUCTION

A time series represents a collection of values obtained from sequential measurements over time. Time series data mining stems from the desire to reify our natural ability to visualize the *shape* of data. Humans rely on complex schemes in order to perform such tasks. We can actually avoid focusing on small fluctuations in order to derive a notion of *shape* and identify almost instantly similarities between patterns on various time scales. Major time series related tasks include query by content [120], anomaly detection [375], motif discovery [232], prediction [374], clustering [229], classification [21] and segmentation [199]. Despite the vast body of work devoted to this topic in the early years, [10] noted that *"the research has not been driven so much by actual problems but by an interest in proposing new approaches"*. However, with the ever-growing maturity of time series data mining techniques, this statement seems to have become obsolete. Nowadays, time series analysis covers a wide range of real-life problems in various fields of research. Some examples include economic forecasting [334], intrusion detection [409], gene expression analysis [234], medical surveillance [63] and hydrology [263].

Time series data mining unveils numerous facets of complexity. The most prominent problems arise from the high dimensionality of time series data and the difficulty of defining a form of similarity measure based on human perception. With the rapid growth of digital sources of information, time series mining algorithms will have to match increasingly massive datasets. These constraints show us that three major issues are involved:

- *Data representation:* How can the fundamental *shape characteristics* of a time series be represented? What invariance properties should the representation satisfy? A representation technique should derive the notion of shape by reducing the dimensionality of data while retaining its essential characteristics.

- *Similarity measurement:* How can any pair of time series be distinguished or matched? How can an intuitive distance between two series be formalized? This measure should establish a notion of similarity based on perceptual criteria, thus allowing the recognition of perceptually similar objects even though they are not mathematically identical.
- *Indexing method:* How should a massive set of time series be organized to enable fast querying? In other words, what *indexing mechanism* should be applied? The indexing technique should provide minimal space consumption and computational complexity.

These implementation components represent the core aspects of time series data mining systems. However these are not exhaustive as many tasks will require the use of more specific modules. Moreover, some of these are useless for some specific tasks. Forecasting (cf. section 5.3.5) is the most blatant example of a topic that requires more advanced analysis processes as it is more closely related to statistical analysis. It may require the use of a time series representation and a notion of similarity (mostly used to measure prediction accuracy) whereas model selection and statistical learning are also at the core of forecasting systems. The components that are *common* to most time series mining tasks have therefore been analyzed and other components found in related tasks have been briefly discussed.

## 5.2 DEFINITIONS

The purpose of this section is to provide a definition for the terms used throughout our study.

**Definition 1.** A *time series* $T$ is an ordered sequence of $n$ real-valued variables

$$T = (t_1, \ldots, t_n), t_i \in \mathbb{R} \tag{5.1}$$

A time series is often the result of the observation of an underlying process in the course of which values are collected from measurements made at uniformly spaced *time instants* and according to a given *sampling rate*. A time series can thus be defined as a set of contiguous time instants. The series can be *univariate* as in definition 1 or *multivariate* when several series simultaneously span multiple dimensions within the same time range.

Time series can cover the full set of data provided by the observation of a process and may be of considerable length. In the case of streaming, they are semi-infinite as time instants continuously feed the series. It thus becomes interesting to consider only the *subsequences* of a series.

**Definition 2.** Given a time series $T = (t_1, \ldots, t_n)$ of length $n$, a *subsequence* $S$ of $T$ is a series of length $m \leqslant n$ consisting of contiguous time instants from $T$

$$S = (t_k, t_{k+1}, \ldots, t_{k+m-1}) \tag{5.2}$$

with $1 \leqslant k \leqslant n - m + 1$. We denote the set of all subsequences of length $m$ from $T$ as $\mathbf{S}_T^m$.

For easier storage, massive time series sets are usually organized in a database.

**Definition 3.** A *time series database* $DB$ is an unordered set of time series.

As one of the major issues with time series data mining is the *high dimensionality* of data, the database usually contains only simplified representations of the series.

**Definition 4.** Given a time series $T = (t_1, ..., t_n)$ of length $n$, a *representation* of $T$ is a model $\bar{T}$ of reduced dimensionality $\bar{d}$ ($\bar{d} \ll n$) such that $\bar{T}$ closely approximates $T$.

Nearly every task of time series data mining relies on a notion of similarity between series. After defining the general principle of similarity measures between time series, we will see (section 5.4.3) how these can be specified.

**Definition 5.** The *similarity measure* $\mathcal{D}(T, U)$ between time series $T$ and $U$ is a function taking two time series as inputs and returning a *distance* $d$ between these series.

This distance has to be *non-negative*, i.e. $\mathcal{D}(T, U) \geqslant 0$. If this measure satisfies the additional *symmetry* property $\mathcal{D}(T, U) = \mathcal{D}(U, T)$ and *subadditivity* $\mathcal{D}(T, V) \leqslant \mathcal{D}(T, U) + \mathcal{D}(U, V)$ (also known as the *triangle inequality*), the distance is said to be a *metric*. As will be seen below (section 5.4.4), on the basis of the triangle inequality, metrics are very efficient measures for indexing. We may also extend this notion of distance to the subsequences.

**Definition 6.** The *subsequence similarity measure* $\mathcal{D}_{subseq}(T, S)$ is defined as

$$\mathcal{D}_{subseq}(T, S) = \min\left(\mathcal{D}\left(T, S'\right)\right) \tag{5.3}$$

for $S' \in \mathbf{S}_S^{|T|}$. It represents the distance between $T$ and its best matching location in $S$.

## 5.3 TASKS IN TIME SERIES DATA MINING

This section provides an overview of the tasks that have attracted wide research interest in time series data mining. These tasks are usually just defined as theoretical objectives though concrete applications may call for simultaneous use of multiple tasks.

### 5.3.1 *Query by content*

Query by content is the most active area of research in time series analysis. It is based on retrieving a set of solutions that are most similar to a query provided by the user. Figure 7 depicts a typical query by content task, represented on a 2-dimensional search space. We can define it formally as

**Definition 7.** *Query by content* - Given a query time series $Q = (q_1, ..., q_n)$ and a similarity measure $\mathcal{D}(Q, T)$, find the ordered list $\mathcal{L} = \{T_1, \ldots, T_n\}$ of time series in the database DB, such that $\forall T_k, T_j \in \mathcal{L}$, $k > j \Leftrightarrow \mathcal{D}(Q, T_k) > \mathcal{D}\left(Q, T_j\right)$.

The content of the result set depends on the *type* of query performed over the database. The previous definition is in fact a generalized formalization of a query by content. It is possible to specify a threshold $\epsilon$ and retrieve all series whose similarity with the query $\mathcal{D}(Q, T)$ is less than $\epsilon$. This type of query is called an *$\epsilon$-range query*.

**Definition 8.** $\epsilon$-range query - Given a query time series $Q = (q_1, ..., q_n)$, a time series database DB, a similarity measure $\mathcal{D}(Q, T)$ and a threshold $\epsilon$, find the set of series $\mathcal{S} = \{T_i \mid T_i \in DB\}$ that are within distance $\epsilon$ from Q. More precisely, find $\mathcal{S} = \{T_i \in DB \mid \mathcal{D}(Q, T_i) \leqslant \epsilon\}$

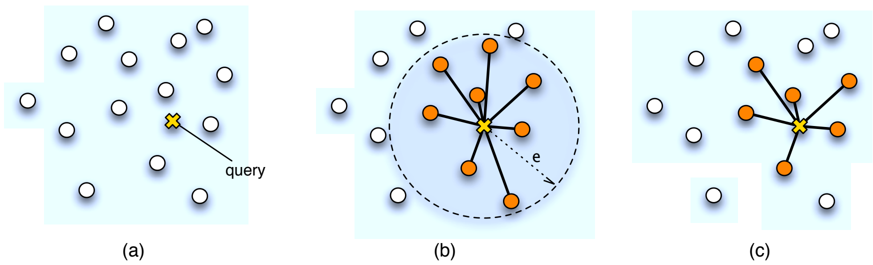



Figure 7: Diagram of a typical query by content task represented in a 2-dimensional search space. Each point in this space represents a series whose coordinates are associated with its features. (a) When a query is entered into the system, it is first transformed into the same representation as that used for other datapoints. Two types of query can then be computed. (b) A $\epsilon-$*range query* will return the set of series that are within distance $\epsilon$ of the query. (c) A $K-$*Nearest Neighbors query* will return the $K$ points closest to the query.

Selecting this threshold is obviously highly data-dependent. Users usually want to retrieve a set of solutions by constraining the number of series it should contain, without knowing how far they will be from the query. It is thus possible to query the $K$ most similar series in the database ($K-$Nearest Neighbors query).

**Definition 9.** *K-Nearest Neighbors* - Given a query time series $Q = (q_1, ..., q_n)$, a time series database $DB$, a similarity measure $\mathcal{D}(Q, T)$ and an integer $K$, find the set of $K$ series that are the most similar to $Q$. More precisely, find $\mathcal{S} = \{T_i \mid T_i \in DB\}$ such that $|\mathcal{S}| = K$ and $\forall T_j \notin \mathcal{S}$, $\mathcal{D}(Q, T_i) \leqslant \mathcal{D}\left(Q, T_j\right)$

Such queries can be called on complete time series; however, the user may also be interested in finding every subsequence of the series matching the query, thus making a distinction between *whole series matching* and *subsequence matching*.

**Definition 10.** *Whole series matching* - Given a query $Q$, a similarity measure $\mathcal{D}(Q, T)$ and a time series database $DB$, find all series $T_i \in DB$ such that $\mathcal{D}(Q, T_i) \leqslant \epsilon$

The distinction between these types of queries is here expressed in terms of $\epsilon-$range query

**Definition 11.** *Subsequence matching* - Given a query $Q$, a similarity measure $\mathcal{D}(Q, T)$ and a database $DB$, find all subsequences $T_i^{'}$ of $T_i \in DB$ such that $\mathcal{D}_{subseq}\left(Q, T_i^{'}\right) \leqslant \epsilon$

In former times, time series mining was almost exclusively devoted to this task (cf. seminal work by [5]). In this paper, the representation was based on a set of coefficients obtained from a Discrete Fourier Transform (DFT) to reduce the dimensionality of data. These coefficients were then indexed with a R*-tree [30]. False hits were removed in a post-processing step, applying the Euclidean distance to complete time series. This paper laid the foundations of a reference framework that many subsequent works just enlarged by using properties of the DFT [294] or similar decompositions such as Discrete Wavelet Transform (DWT) [81], that has been shown to have similar efficiency depending on the dataset at hand [284]. The Discrete Cosine Transform (DCT) has also been suggested [210] but it appeared later that it did not have any advantage over other decompositions [201]. Several numeric transformations – such as random projections [176], Piecewise Linear Approximation (PLA) [327], Piecewise Approximate Aggregation (PAA) [198, 395] and Adaptive Piecewise Constant Approximation (APCA) [197] – have been used as representations. Symbolic representations have also been widely used. A shape alphabet with fixed resolution was originally proposed in [6]. Other symbolic representations have been proposed, such as the bit level approximation [297] or the Symbolic Aggregate approXimation (SAX) [231]; the latter one has been shown to outperform most of the other representations [337]. We will find below a detailed overview of representations (section 5.4.2), distance measures (section 5.4.3) and indexing techniques (section 5.4.4).

Other important extensions to query by content include the handling of scaling and gaps [364], noise [366], query constraints [141] and time warping, either by allowing false dismissals [396] or working without constraints [311]. Lower bounding distances without false dismissals for DTW were proposed in [205] and [196] which allows exact indexing. The recent trend of query by content systems seems to be focused on streams. Given the continuously growing bandwidth, most of next generation analysis will most likely have to be performed over stream data. The dynamic nature of streaming time series precludes using the methods proposed for the static case. In a recent study, [209] introduced the most important issues concerning similarity search in static and streaming time series databases. In Kontaki et al. [208], the use of an incremental computation of DFT allows to adapt to the stream update frequency. However, maintaining the indexing tree for the whole streaming series seems to be uselessly costly. Assent et al. [13] proposed a filter-and-refine DTW algorithm called Anticipatory DTW, which allows faster rejection of false candidates. Lian et al. [226] proposed a weighted locality-sensitive hashing (WLSH) technique applying to approximate queries and working by incremental updating adaptive to the characteristics of stream data. Lian and Chen [225] proposed three approaches, polynomial, DFT and probabilistic, to predict future unknown values and answer queries based on the predicated data. This approach is a combination of prediction (cf. section 5.3.5) and streaming query by content; it is representative of an effort to obtain a convergence of approaches that seem to be heterogeneous.

### 5.3.2 *Clustering*

Clustering is the process of finding natural groups, called *clusters*, in a dataset. The objective is to find the most homogeneous clusters that are as distinct as possible from other clusters. More formally, the grouping should maximize inter-cluster variance while minimizing intra-cluster variance. The algorithm should thus automatically locate which groups are intrinsically present in the data. Figure **8** depicts some possible outputs of a clustering algorithm. It can be seen in this figure that the main difficulty concerning any clustering problem (even out of the scope of time series mining) usually lies in defining the correct number of clusters. The time series clustering task can be divided into two sub-tasks.

#### *Whole series clustering*

Clustering can be applied to each complete time series in a set. The goal is thus to regroup entire time series into clusters so that the time series are as similar to each other as possible within each cluster.

**Definition 12.** Given a time series database $DB$ and a similarity measure $\mathcal{D}(Q,T)$, find the set of clusters $\mathcal{C} = \{c_i\}$ where $c_i = \{T_k \mid T_k \in DB\}$ that maximizes inter-cluster distance and minimizes intra-cluster variance. More formally $\forall i_1, i_2, j$ such that $T_{i_1}, T_{i_2} \in c_i$ and $T_j \in c_j$ $\mathcal{D}\left(T_{i_1}, T_j\right) \gg \mathcal{D}\left(T_{i_1}, T_{i_2}\right)$

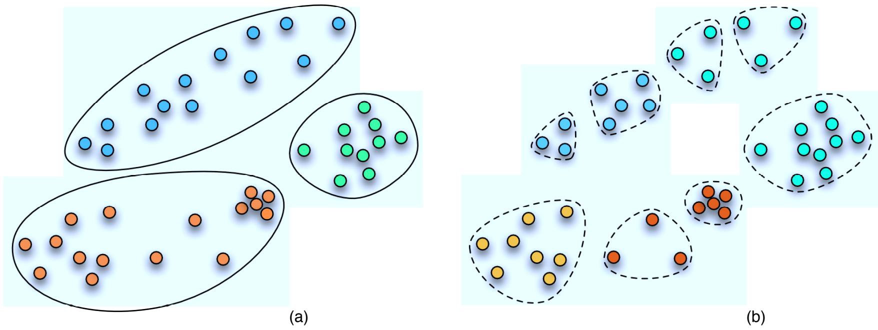



Figure 8: Two possible outputs from the same clustering system obtained by changing the required number of clusters with (a) $N = 3$ and (b) $N = 8$. As we can see, the clustering task is a non trivial problem that highly depends on the way parameters are initialized and the level of detail targeted. This parameter selection issue is common to every clustering task, even out of the scope of time series mining.

There have been numerous approaches for whole series clustering. Typically, after defining an adequate distance function, it is possible to adapt any algorithm provided by the generic clustering topic. Clustering is traditionnally performed by using Self Organizing Maps (SOM) [84], Hidden Markov Models (HMM) [333] or Support Vector Machines (SVM) [397]. [134] proposed a variation of the Expectation Maximization (EM) algorithm. However, this model-based approach has usually some scalability problems and implicitly presupposes the existence of an underlying model which is not straightforward for every dataset. Using Markov chain Monte Carlo (MCMC) methods, Frohwirth and Kaufmann [130] make an estimation about the appropriate grouping of time series simultaneously along with the group-specific model parameters. A good survey of generic clustering algorithms from a data mining perspective is given in Berkin [38]. This review focuses on methods based on classical techniques that can further be applied to time series. A classification of clustering methods for various static data is proposed in Han and Kamber [159] following five categories: *partitioning*, *hierarchical*, *density-based*, *grid-based* and *model-based*. For the specificities of time series data, three of these five categories (partitioning, hierarchical and model-based) have been applied [227]. Clustering of time series is especially useful for data streams; it has been implemented by using clipped data representations [17], Auto-Regressive (AR) models [96], $k-$Means [365] and – with greater efficiency – k-center clustering [97]. Interested readers may refer to [227] who provides a thorough survey of time series clustering issues by discussing the advantages and limitations of existing works as well as avenues for research and applications.

*Subsequence clustering*

In this approach, the clusters are created by extracting subsequences from a single or multiple longer time series.

**Definition 13.** Given a time series $T = (t_1, ..., t_n)$ and a similarity measure $\mathcal{D}(Q, C)$, find the set of clusters $\mathcal{C} = \{c_i\}$ where $c_i = \left\{ T_j' \mid T_j' \in \mathbf{S}_T^n \right\}$ is a set of subsequences that maximizes inter-cluster distance and intra-cluster cohesion.

In [162], the series are sliced into non-overlapping windows. Their width is chosen by investigating the periodical structure of the time series by means of a DFT analysis. This

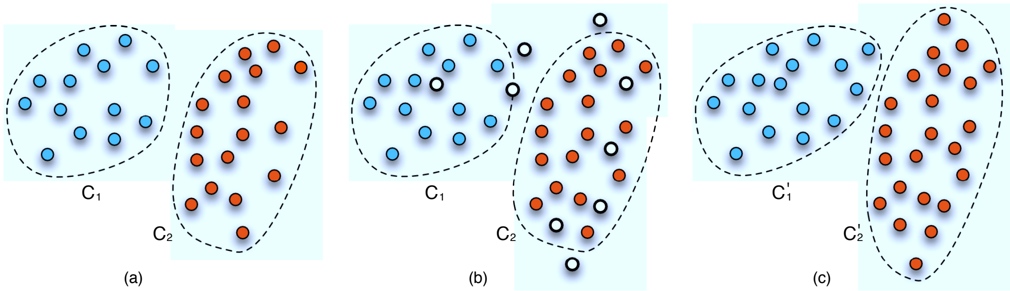



Figure 9: The three main steps of a classification task. (a) A training set consisting of two pre-labeled classes $C_1$ and $C_2$ is entered into the system. The algorithm will first try to learn what the characteristic features distinguishing one class from another are; they are represented here by the class boundaries. (b) An unlabeled dataset is entered into the system that will then try to automatically deduce which class each datapoint belongs to. (c) Each point in the set entered has been assigned to a class. The system can then optionally adapt the classes boundaries.

approach is limited by the fact that, when no strong periodical structure is present in the series, non-overlapping slicing may miss important structures. A straightforward way to extend this approach can therefore be to extract shorter overlapping subsequences and then cluster the resulting set. However, this overlapping approach has been shown to produce meaningless results [200]. Despite these deceptive results, the authors pointed out that a meaningful subsequence clustering system could be constructed on top of a motif mining [273] algorithm (cf. section 5.3.7). Denton [104] was first to suggest an approach to overcome this inconsistency by not forcing the algorithm to use all subsequences in the clustering process. In the context of intrusion detection, Zhong et al. [409] studied multiple centroid-based unsupervised clustering algorithms, and proposed a self-labeling heuristic to detect any attack within network traffic data. Clustering is also one of the major challenges in bioinformatics, especially in DNA analysis. Kerr et al. [203] surveyed state-of-the-art applications of gene expression clustering and provided a framework for the evaluation of results.

### 5.3.3 *Classification*

The classification task seeks to assign labels to each series of a set. The main difference when compared to the clustering task is that classes are known in advance and the algorithm is trained on an example dataset. The goal is first to learn what the distinctive *features* distinguishing classes from each others are. Then, when an unlabeled dataset is entered into the system, it can automatically determine which class each series belongs to. Figure 9 depicts the main steps of a classification task.

**Definition 14.** Given an unlabeled time series $T$, assign it to one class $c_i$ from a set $\mathcal{C} = \{c_i\}$ of predefined classes.

There are two types of classification. The first one is the *time series classification* similar to whole series clustering. Given sets of time series with a label for each set, the task consists in training a classifier and labeling new time series. An early approach to time series classification was presented in [21]. However, it is based on simple trends whose results are therefore hard to interpret. A piecewise representation was later proposed in [195], it is robust to noise and weighting can be applied in a relevance feedback framework. The same representation was used in [136]; it is apparently not too robust to outliers. To overcome the obstacle of high dimensionality, [185] used Singular Value Decomposition to select essential frequencies. However, it implies higher computational costs. In a recent study, [307] compared three types of classifiers: nearest neighbor, support vector machines and decision forests. All three methods seems to be valid, though highly depending on the dataset at hand. 1-NN classification algorithm with DTW seems to be the most widely used classifier; it was shown to be highly accurate [382], though computing speed is significantly affected by repeated DTW computations. To overcome this limitation [336] proposed a template construction algorithm based on the Accurate Shape Averaging (ASA) technique. Each training class is represented by only one sequence so that any incoming series is compared only with one averaged template per class. Several other techniques have been introduced, such as ARMA models [103] or HMM [408]. In the context of clinical studies, [234] enhanced HMM approaches by using discriminative HMMs in order to maximize inter-classes differences. Using the probabilistic transitions between fewer states results in the patients being aligned to the model and can account for varying rates of progress. This approach has been applied in [237], in order to detect post-myocardial infarct patients. Several machine learning techniques have also been introduced such as neural networks [260] or Bayesian classification [285]. However, many of these proposals have been shown to be overpowered by a simple 1NN-DTW classifier [382]. A double-loop EM algorithm with a Mixture of Experts network structure has been introduced in [340] for the detection of epileptic seizure based on the EEG signals displayed by normal and epileptic patients. A well-known problem in classification tasks is the *overtraining*, i.e. when too many training data lead to an over-specified and inefficient model. [296] suggested a stopping criterion to improve the data selection during a self training phase. [405] proposed a time series reduction, which extracts patterns that can be used as inputs to classical machine-learning algorithms. Many interesting applications to this problem have been investigated such as brain-computer interface based on EEG signals; they have been reviewed in [236].

### 5.3.4 *Segmentation*

The segmentation (or *summarization*) task aims at creating an accurate approximation of time series, by reducing its dimensionality while retaining its essential features. Figure 10 shows the output of a segmentation system. Section 5.4.2 will show that most time series representations try to solve this problem implicitly. We can state the segmentation problem more formally as

**Definition 15.** Given a time series $T = (t_1, ..., t_n)$, construct a model $\bar{T}$ of reduced dimensionality $\bar{d}$ ($\bar{d} \ll n$) such that $\bar{T}$ closely approximates $T$. More formally $\left|R\left(\bar{T}\right) - T\right| < \epsilon_r$, $R\left(\bar{T}\right)$ being the reconstruction function and $\epsilon_r$ an error threshold.

The objective of this task is thus to minimize the reconstruction error between a reduced representation and the original time series. The main approach that have been undertaken over the years seems to be Piecewise Linear Approximation (PLA) [327]. The main idea behind PLA is to split the series into most representative segments, and then fit a polynomial model for each segment. A good review on the most common segmentation methods in the context of PLA representation can be found in [199]. Three basic approaches are distinguished. In *sliding windows*, a segment is grown until it exceeds some error threshold [327]. This approach has shown poor performance with many real life datasets [199]. The *top-down* approach consists in recursively

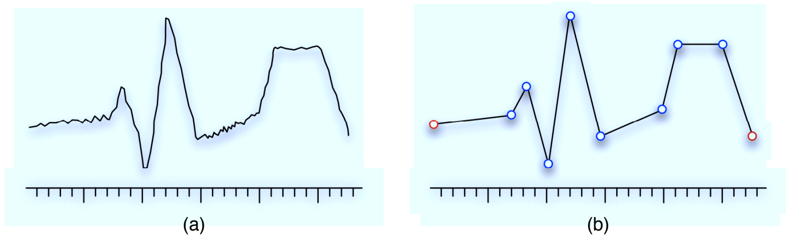



Figure 10: Example of application of a segmentation system. From (a) usually noisy time series containing a very large number of datapoints, the goal is to find (b) the closest approximation of the input time series with the maximal dimensionality reduction factor without loosing any of its essential features.

partitioning a time series until some stopping criterion is met [221]. This approach has time complexity $O\left(n^2\right)$ [271] and is qualitatively outperformed by *bottom-up*. In this approach, starting from the finest approximation, segments are iteratively merged [195]. [169] present fast greedy algorithms to improve previous approaches and a statistical method for choosing the number of segments is described in [362].

Several other methods have been introduced to handle this task. [266] introduced a representation of time series that implicitly handles the segmentation of time series. They proposed user-specified amnesic functions reducing the confidence to older data in order to make room for newer data. In the context of segmenting hydrological time series, [192] proposed a maximum likelihood method using an HMM algorithm. However, this method offers no guarantee to yield the globally optimal segmentation without long execution times. For dynamic summary generation, [261] proposed an online transform-based summarization techniques over data streams that can be updated continuously. The segmentation of time-series can also be seen as a constrained clustering problem. [2] proposed to group time points by their similarity, provided that all points in a cluster come from contiguous time instants. Therefore, each cluster represents the segments in time whose homogeneity is evaluated with a local PCA model.

### 5.3.5 *Prediction*

In several fields of research, time series are very long and considered *smooth*, their subsequent values are within predictable ranges of one another [326]. The task of prediction is aimed at explicitly modeling such variable dependencies to forecast the next few values of a series. Figure 11 depicts various forecasting scenarios.

**Definition 16.** Given a time series $T = (t_1, ... t_n)$, predict the $k$ next values $(t_{n+1}, ..., t_{n+k})$ that are most likely to occur.

Prediction is a major area in several fields of research. Concerning time series, it is one of the most extensively applied tasks. Literature about this is so abundant that dozens of reviews can focus on only a specific field of application or family of learning methods. Even if it can use time series representations and a notion of similarity to evaluate accuracy, It also relies on several statistical components that are out of the scope of this article, e.g. model selection and statistical learning. This task will be mentioned because of its importance but the interested reader willing to have further information may consult several references on forecasting [58, 160, 356, 59] Several methods have been applied to this task. A natural option could be AR models

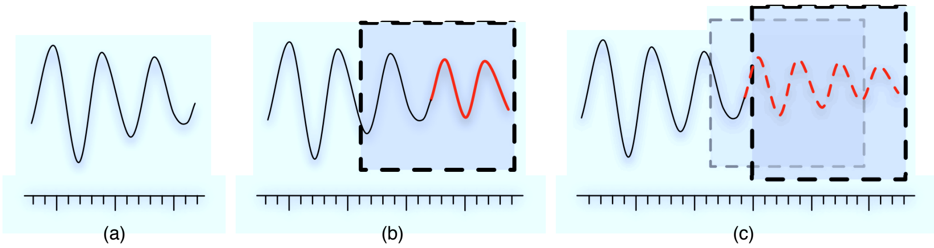



Figure 11: A typical example of the time series prediction task. (a) The input time series may exhibit a periodical and thus predictable structure. (b) The goal is to forecast a maximum number of upcoming datapoints within a prediction window. (c) The task becomes really hard when it comes to having *recursive prediction*, i.e. the long term prediction of a time series implies reusing the earlier forecast values as inputs in order to go on predicting.

[55]. These models have been applied for a long time to prediction tasks involving signal de-noising or dynamic systems modeling. It is however possible to use more complex approaches such as neural networks [211] or clusters function approximation [324] to solve this problem. A polynomial architecture has been developed to improve a multilayer neural network in [388] by reducing higher-order terms to a simple product of linear functions. Other learning algorithms, such as SOM, provided efficient supervised architectures. A survey of applications of SOM to time series prediction is given in [26]. Recent improvements for time series forecasting have been proposed; [279] proposed a Bayesian prediction for time series subject to discrete breaks, handling the size and duration of possible breaks by means of a hierarchical HMM. A dynamic genetic programming (GP) model tailored for forecasting streams was proposed in [370] by adapting incrementally based on retained knowledge. The prediction task seems one of the most commonly applied in real-life applications, considering that market behavior forecasting relies on a wealth of financial data. Bai and Ng [20] proposed to refine the method of factor forecasting by introducing 'targeted predictors' selected by using a hysteresis (hard and soft thresholding) mechanism. The prediction task has also a wide scope of applications ranging from tourism demand forecasting [334] to medical surveillance [63]. In this paper, the authors compared the predictive accuracy of three methods, namely, non-adaptive regression, adaptive regression, and the Holt-Winters method; the latter appeared to be the best method. In a recent study, Ahmed et al. [7] carried out a large scale comparison for the major machine-learning models applied to time series forecasting, following which the best two methods turned out to be multilayer perceptron and Gaussian process regression. However, learning a model for long-term prediction seems to be more complicated, as it can use its own outputs as future inputs (*recursive prediction*). [168] proposed the use of least-squares SVM, to solve this problem. [69] further applied saliency analysis to SVM in order to remove irrelevant features based on the sensitivity of the network output to the derivative of the feature input. [335] proposed to combine direct prediction and an input selection in order to cope with long-term prediction of time series.

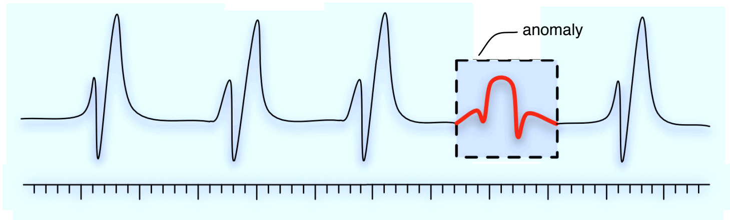



Figure 12: An idealized example of the anomaly detection task. A long time series which exhibits some kind of periodical structure can be modeled thanks to a reduced pattern of "standard" behavior. The goal is thus to find subsequences which does not follow the model and may therefore be considered as anomalies.

### 5.3.6 *Anomaly detection*

The detection of anomalies seeks to find abnormal subsequences in a series. Figure 12 depicts an example of anomaly detection. It has numerous applications ranging from biosurveillance [93] to intrusion detection [409].

**Definition 17.** Given a time series $T = (t_1, ..., t_n)$ and a model of its normal behavior, find all subsequences $T' \in \mathbf{S}_T^n$ which contain anomalies, i.e. do not fit the model.

A good discussion on the difficulties of mining rare events is given in [375]. The usual approach to detect anomalies is to first create a model of a series' normal behavior and characterize subsequences that stray too far from the model as anomalies. This approach can be linked to the prediction task. Indeed, if we can forecast the next values of a time series with good accuracy, outliers can be detected in a straightforward manner and flagged as anomalies. This approach was undertaken first in [399] using SOM model to represent the expected behavior. A framework for novelty detection is defined in [239] and implemented based on Support Vector Regression (SVR). Machine learning techniques were also introduced to dynamically adapt their model of normal behavior. [8] investigated the use of block-based One-Class Neighbor Machine and recursive Kernel-based algorithms and showed their applicability to anomaly detection. [88] proposed two algorithms to find anomalies in the Haar wavelet coefficients of the time series. A state-based approach is taken in [313] using time point clustering so that clusters represents the normal behavior of a series. Another definition of anomalies, the time series *discords*, are defined as subsequences that are maximally different from all the remaining subsequences [202]. This definition is able to capture the idea of most unusual subsequence within a time series and its unique parameter is the required length of the subsequences. Thanks to this definition [391] proposed an exact algorithm that requires only two linear scans, thus allowing for the use of massive datasets. However, as several proposals, the number of anomalous subsequences must be specified prior to the search. Several real-life applications have also been outlined in recent research. Anomaly detection is applied in [155] to detect fatigue damage in polycrystalline alloys, thus preventing problems in mechanical structures. An anomaly detection scheme for time series is used in [93] to determine whether streams coming from sensors contain any abnormal heartbeats. A recent overview and classification of the research on anomaly detection is presented in [82], which provides a discussion on the computational complexity of each technique.

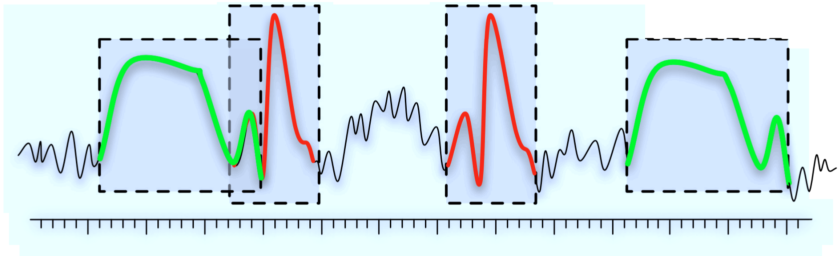


Figure 13: The task of motif discovery consists in finding every subsequence that appears recurrently in a longer time series. These subsequences are named motifs. This task exhibits a high combinatorial complexity as several motifs can exist within a single series, motifs can be of various lengths and even overlap.

### 5.3.7 *Motif discovery*

Motif discovery consists in finding every subsequences (named *motif*) that appears recurrently in a longer time series. This idea was transferred from gene analysis in bioinformatics. Figure 13 depicts a typical example of motif discovery. Motifs were defined originally in [273] as *typical* non-overlapping subsequences. More formally

**Definition 18.** Given a time series $T = (t_1, \ldots, t_n)$, find all subsequences $T' \in \mathbf{S}_T^n$ that occurs repeatedly in the original time series.

A great interest for this research topic has been triggered by the observation that subsequence clustering produces meaningless results [200]. The authors pointed out that motif discovery could be used as a subroutine to find meaningful clusters. In order to find motifs more efficiently, [92] proposed to use the random projection algorithm [62] which was successfully used for DNA sequences. However, because of its probabilistic nature, it is not guaranteed to find the exact set of motifs. [122] proposed an algorithm that can extract approximate motifs in order to mine time series data from protein folding/unfolding simulations. In [235], motif discovery is formalized as a continuous top-k motif balls problem in an m-dimensional space. However, the efficiency of this algorithm critically depends on setting the desired length of the pattern. [349] introduced a k-motif-based algorithm that provides an interesting mechanism to generate summaries of motifs. [390] showed that motif discovery can be severely altered by any slight change of *uniform scaling* (linear stretching of the pattern length) and introduced a scaling-invariant algorithm to determine the motifs. An algorithm for exact discovery of time series motifs has been recently proposed [258], which is able to process very large datasets by using early abandoning on a linear re-ordering of data. [255] studied the constrained motif discovery problem which provides a way to incorporate prior knowledge into the motif discovery process. They showed that most unconstrained motif discovery problems can be transformed into constrained ones and provided two algorithms to solve such problem. The notion of motifs can be applied to many different tasks. The modeling of normal behavior for anomaly detection (cf. section 5.3.6) implies finding the recurrent motif of a series. For time series classification, significant speed-ups can be achieved by constructing motifs for each class [405].

## 5.4 IMPLEMENTATION COMPONENTS

In this section, we review the implementation components common to most of time series mining tasks. As said earlier, the three key aspects when managing time series data are *representation* methods, *similarity* measures and *indexing* techniques. Because of the *high dimensionality* of time series, it is crucial to design low-dimensional representations that preserve the fundamental characteristics of a series. Given this representation scheme, the *distance* between time series needs to be carefully defined in order to exhibit perceptually relevant aspects of the underlying similarity. Finally the indexing scheme must allow to efficiently manage and query evergrowing massive datasets.

### 5.4.1 *Preprocessing*

In real-life scenarios, time series usually come from live observations [302] or sensors [337] which are particularly subject to noise and outliers. These problems are usually handled by preprocessing the data. Noise filtering can be handled by using traditional signal processing techniques like digital filters or wavelet thresholding. In [170], Independent Component Analysis (ICA) is used to extract the main mode of the series. As will be explained in section 5.4.2, several representations implicitly handle noise as part of the transformation.

The second issue concerns the scaling differences between time series. This problem can be overcome by a linear transformation of the amplitudes [141]. Normalizing to a fixed range [6] or first subtracting the mean (known as *zero mean / unit variance* [197]) may be applied to both time series, however it does not give the optimal match of two series under linear transformations [11]. In [142] the transformation is sought with optional bounds on the amount of scaling and shifting. However, normalization should be handled with care. As noted by [364], normalizing an essentially flat but noisy series to unit variance will completely modify its nature and normalizing small enough subsequences can provoke all series to look the same [229].

Finally, resampling (or *uniform time warping* [264]) can be performed in order to obtain series of the same length [194]. Down-sampling the longer series has been shown to be fast and robust [11].

### 5.4.2 *Representation*

As mentioned earlier, time series are essentially high dimensional data. Defining algorithms that work directly on the raw time series would therefore be computationally too expensive. The main motivation of representations is thus to emphasize the essential characteristics of the data in a concise way. Additional benefits gained are efficient storage, speedup of processing as well as implicit noise removal. These basic properties lead to the following requirements for any representation:

- Significant reduction of the data dimensionality
- Emphasis on fundamental shape characteristics on both *local* and *global* scales
- Low computational cost for computing the representation
- Good reconstruction quality from the reduced representation
- Insensitivity to noise or implicit noise handling

Many representation techniques have been investigated, each of them offering different trade-offs between the properties listed above. It is however possible to classify these approaches according to the kind of transformations applied. In order to perform such classification, we follow the taxonomy of [201] by dividing representations into three categories, namely *non data-adaptive*, *data-adaptive* and *model-based*. Figure 14 synthesize every reviewed representation based on our classification.

### *Non Data-Adaptive*

In non data-adaptive representations, the parameters of the transformation remain the same for every time series regardless of its nature. The first non data-adaptive representations were drawn from spectral decompositions. The DFT was used in the seminal work of [5]. It projects the time series on a sine and cosine functions basis [120] in the real domain. The resulting representation is a set of sinusoidal coefficients. Instead of using a fixed set of basis functions, the DWT uses scaled and shifted versions of a mother *wavelet* function [81]. This gives a multi-resolution decomposition where low frequencies are measured over larger intervals thus providing better accuracy [284]. A large number of wavelet functions have been used in the literature like Haar [80], Daubechies [284] or Coiflets [326]. The Discrete Cosine Transform (DCT) uses only a cosine basis; it has also been applied to time series mining [210]. However, it has been shown that it does not offer any advantage over previously cited decompositions [201]. Finally, an approximation by Chebychev polynomials [65] has also been proposed but the results obtained have later been withdrawn due to an error in implementation.

Other approaches – more specific to time series – have been proposed. The Piecewise Aggregate Approximation (PAA) introduced by [198] (submitted independently as Segmented Means in [395]) represents a series through the mean values of consecutive fixed-length segments. An extension of PAA including a multi-resolution property (MPAA) has been proposed in [229]. [15] suggested to extract a sequence of amplitude-levelwise local features (ALF) to represent the characteristics of local structures. It was shown that this proposal provided weak results in [106]. Random projections have been used for representation in [176]; in this case, each time series enters a convolution product with k random vectors drawn from a multivariate standard. This approach has recently been combined with spectral decompositions by [302] with the purpose of answering statistical queries over streams.

### *Data-Adaptive*

This approach implies that the parameters of a transformation are modified depending on the data available. By adding a data-sensitive selection step, almost all non data-adaptive methods can become data-adaptive. For spectral decompositions, it usually consists in selecting a subset of the coefficients. This approach has been applied to DFT [366] and DWT [339]. A data-adaptive version of PAA has been proposed in [246], with vector quantization being used to create a codebook of recurrent subsequences. This idea has been adapted to allow for multiple resolution levels [247]. However, this approach has only been tested on smaller datasets. A similar approach has been undertaken in [337] with a codebook based on motion vectors being created to spot gestures. However, it has been shown to be computationally less efficient than SAX.

Several inherently data-adaptive representations have also been used. SVD has been proposed [210] and later been enhanced for streams [301]. However, SVD requires computation of eigenvalues for large matrices and is therefore far more expensive than

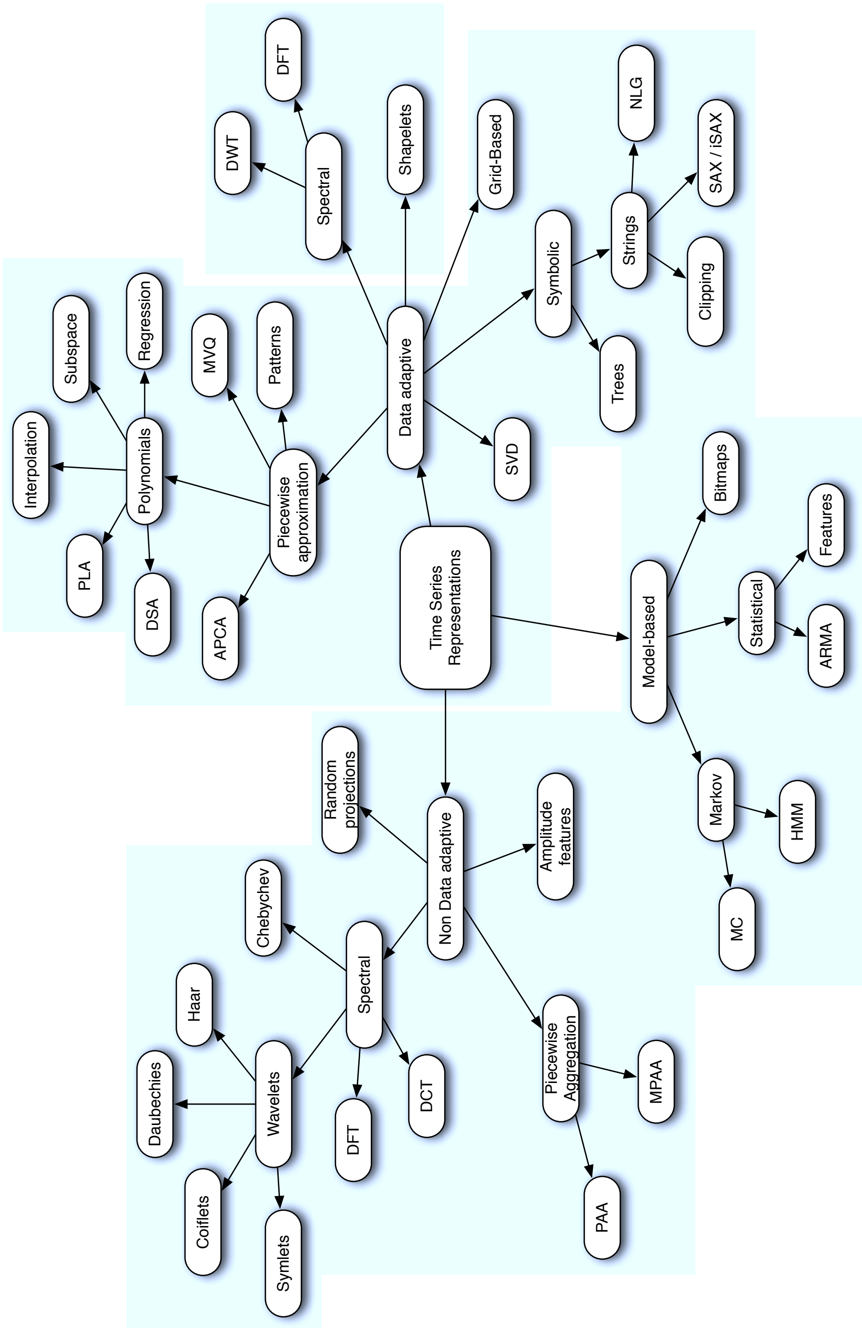



Figure 14: Classification of the time series representations reviewed in this chapter.

other schemes mentioned. It has recently been adapted to find multi-scale patterns in time series streams [268]. PLA [327] is a widely used approach for the segmentation task (cf. section 5.3.4) The set of polynomial coefficients can be obtained either by interpolation [195] or regression [172]. Many derivatives of this technique have been introduced. The Landmarks system [277] extends this notion to include a multi-resolution property. However, the extraction of features relies on several parameters which are highly data-dependent. APCA [197] uses constant approximations per segment instead of polynomial fitting. Indexable PLA has been proposed by [87] to speed up the indexing process. [265] put forward an approach based on PLA, to answer queries about the recent past with greater precision than older data and called such representations amnesic. The method consisting in using a segmentation algorithm as a representational tool has been extensively investigated. The underlying idea is that segmenting a time series can be equated with the process of representing the most salient features of a series while considerably reducing its dimensionality. [385] proposed a pattern-based representation of time series. The input series is approximated by a set of concave and convex patterns to improve the subsequence matching process. [402] proposed a pattern representation of time series to extract outlier values and noise. The Derivative Segment Approximation (DSA) model [152] is a representation based on time series segmentation through an estimation of derivatives to which DTW can be applied. The polynomial shape space representation [132] is a subspace representation consisting of trend aspects estimators of a time series. [24] put forward a two-level approach to recognize gestures by describing individual trajectories with key-points, then characterizing gestures through the global properties of the trajectories.

Instead of producing a numeric output, it is also possible to discretize the data into symbols. This conversion into a symbolical representation also offers the advantage of implicitly performing noise removal by complexity reduction. A relational tree representation is used in [22]. Non-terminal nodes of the tree correspond to valleys and terminal nodes to peaks in the time series. The Symbolic Aggregate approXimation (SAX) [231], based on the same underlying idea as PAA, calls on equal frequency histograms on sliding windows to create a sequence of short words. An extension of this approach, called indexable Symbolic Aggregate approXimation (iSAX) [328], has been proposed to make fast indexing possible by providing zero overlap at leaf nodes. The grid-based representation [9] places a two dimensional grid over the time series. The final representation is a bit string describing which values were kept and which bins they were in. Another possibility is to discretize the series to a binary string (a technique called *clipping*) [297]. Each bit indicates whether the series is above or below the average. That way, the series can be very efficiently manipulated. In [19] this is done using the median as the clipping threshold. Clipped series offer the advantage of allowing direct comparison with raw series, thus providing a tighter lower bounding metric. Thanks to a variable run-length encoding, [18] show that it is also possible to define an approximation of the Kolmogorov complexity. Recently, a very interesting approach has been proposed in [394]; it is based on primitives called *shapelets*, i.e. subsequences which are maximally representative of a class and thus fully discriminate classes through the use of a dictionary. This approach can be considered as a step forward towards bridging the gap between time series and shape analysis.

*Model-based*

The model-based approach is based on the assumption that the time series observed has been produced by an underlying model. The goal is thus to find parameters of such a model as a representation. Two time series are therefore considered similar if they have been produced by the same set of parameters driving the underlying model. Several parametric temporal models may be considered, including statistical modeling by feature extraction [260], ARMA models [188] Markov Chains (MCs) [321] or HMM [267]. MCs are obviously simpler than HMM so they fit well shorter series but their expressive power is far more limited. The Time Series bitmaps introduced in [214] can also be considered as a model-based representation for time series, even if it mainly aims at providing a visualization of time series.

### 5.4.3 *Similarity measure*

Almost every time series mining task requires a subtle notion of similarity between series, based on the more intuitive notion of *shape*. When observing simultaneously multiple characteristics of a series, humans can abstract from such problems as amplitude, scaling, temporal warping, noise and outliers. The Euclidean distance is obviously unable to reach such a level of abstraction. Numerous authors have pointed out several pitfalls when using $L_p$ norms [106, 194, 395]. However, it should be noted that, in the case of very large datasets, Euclidean distance has been shown [328] to be sufficient as there is a larger probability that an almost exact match exists in the database. Otherwise, a similarity measure should be consistent with our intuition and provide the following properties:

1. It should provide a recognition of perceptually similar objects, even though they are not mathematically identical;
2. It should be consistent with human intuition;
3. It should emphasize the most salient features on both *local* and *global* scales;
4. A similarity measure should be universal in the sense that it allows to identify or distinguish arbitrary objects, i.e. no restrictions on time series are assumed;
5. It should abstract from distortions and be invariant to a set of transformations.

Many authors have reported about various transformation invariances required for similarity. Given a time series $T = \{t_1, \ldots, t_n\}$ of $n$ datapoints, we consider the following transformations:

- *Amplitude shifting*: The series $G = \{g_1, \ldots, g_n\}$ obtained by a linear amplitude shift of the original series $g_i = t_i + k$ with $k \in \mathbb{R}$ a constant.
- *Uniform amplification*: The series $G$ obtained by multiplying the amplitude of the original series $g_i = k.t_i$ with $k \in \mathbb{R}$ a constant.
- *Uniform time scaling*: The series $G = \{g_1, \ldots, g_m\}$ produced by a uniform change of the time scale of the original series $g_i = t_{\lceil k.i \rceil}$ with $k \in \mathbb{R}$ a constant.
- *Dynamic amplification*: The series $G$ obtained by multiplying the original series by a dynamic amplification function $g_i = h(i).t_i$ with $h(i)$ a function such that $\forall t \in [1 \ldots n]$, $h'(t) = 0$ if and only if $t'_i = 0$.
- *Dynamic time scaling*: The series $G$ obtained by a dynamic change of the time scale $g_i = t_{h(i)}$ with $h(i)$ a positive and strictly increasing function such $h : \mathbb{N} \to [1 \ldots n]$

- *Additive Noise*: The series $G$ obtained by adding a noisy component to the original series $g_i = t_i + \epsilon_i$ with $\epsilon_i$ an independent identically distributed white noise.
- *Outliers*: The series $G$ obtained by adding outliers at random positions. Formally, for a given set of random time positions $\mathcal{P} = \{k \mid k \in [1 \ldots n]\}$, $g_k = \epsilon_k$ with $\epsilon_k$ an independent identically distributed white noise.

The similarity measure $\mathcal{D}(T, G)$ should be robust to any combinations of these transformations. This property lead to our formalization of four general types of robustness.

*Axioms of robustness*

We introduce in this section novel properties expressing robustness for *scaling* (amplitude modifications), *warping* (temporal modifications), *noise* and *outliers*. Let $\mathcal{S}$ be a collection of time series, and let $\mathcal{H}$ be the maximal group of homeomorphisms under which $\mathcal{S}$ is closed. A similarity measure $\mathcal{D}$ on $\mathcal{S}$ is called *scale robust* if it satisfies :

**Proposition.** *For each* $T \in \mathcal{S}$ *and* $\alpha > 0$ *there is a* $\delta > 0$ *such that* $\|t_i - h(t_i)\| < \delta$ *for all* $t_i \in T$ *implies* $\mathcal{D}(T, h(T)) < \alpha$ *for all* $h \in \mathcal{H}$.

We call a similarity measure *warp robust* if the following holds :

**Proposition.** *For each* $T = \{t_i\} \in \mathcal{S}, T' = \left\{t_{h(i)}\right\}$ *and* $\alpha > 0$ *there is a* $\delta > 0$ *such that* $\|i - h(i)\| < \delta$ *for all* $t_i \in T$ *implies that* $\mathcal{D}(T, T') < \alpha$ *for all* $h \in \mathcal{H}$.

We call a similarity measure *noise robust* if it satisfies the following property :

**Proposition.** *For each* $T \in \mathcal{S}$ *and* $\alpha > 0$, *there is a* $\delta > 0$ *such that* $U = T + \epsilon$ *with* $p(\epsilon) = \mathcal{N}(0, \delta)$ *implies* $\mathcal{D}(T, U) < \alpha$ *for all* $U \in \mathcal{S}$

We call a measure *outlier robust* if the following holds :

**Proposition.** *For each* $T \in \mathcal{S}$, $\mathcal{K} = \{\mathrm{rand}[1...n]\}$ *and* $\alpha > 0$, *there is a* $\delta > 0$ *such that if* $|\mathcal{K}| < \delta$ *and* $U_{k \in \mathcal{K}} = \epsilon_k$ *and* $U_{k \notin \mathcal{K}} = T_k$ *implies* $\mathcal{D}(T, U) < \alpha$ *for all* $U \in \mathcal{S}$

Similarity measures can be classified in four categories. *Shape-based* distances compare the overall shape of the series. *Edit-based* distances compare two time series on the basis of the minimum number of operations needed to transform one series into another one. *Feature-based* distances extract features describing aspects of the series that are then compared with any kind of distance function. *Structure-based* similarity aims at finding higher-level structures in the series to compare them on a more global scale. We further subdivide this category into two specific subcategories. *Model-based* distances work by fitting a model to the various series and then comparing the parameters of the underlying models. *Compression-based* distances analyze how well two series can be compressed together. Similarity is reflected by higher compression ratios. As defined by [194], we refer to distance measures that compare the $i$th point of a series to the $i$th point of another as *lock-step* and measures that allow flexible (one-to-many / one-to-none) comparison as *elastic*. Figure 15 synthesizes every reviewed distance measure based on this classification.

*Shape-based*

The Euclidean distance and other $L_p$ norms [395] have been the most widely used distance measures for time series [194]. However, these have been shown to be poor

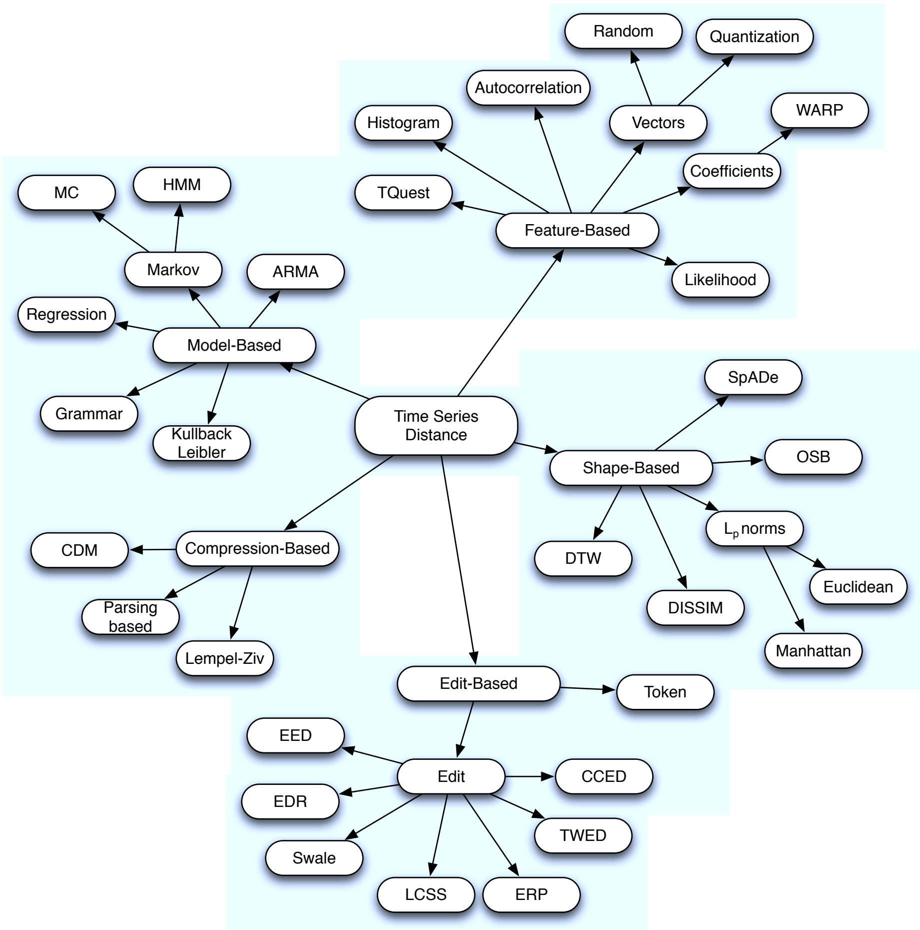



Figure 15: Classification of the distance measures reviewed in this chapter.

similarity measurements [10, 106]. As a matter of fact, these measures does not match any of the types of robustness. Even if the problems of scaling and noise can be handled in a preprocessing step [141], the warping and outliers issues need to be addressed with more sophisticated techniques. This is where the use of elastic measures can provide an elegant solution to both problems.

Handling the local distortions of the time axis is usually addressed using *non-uniform time warping* [195], more specifically with Dynamic Time Warping (DTW) [40]. This measure is able to match various sections of a time series by allowing warping of the time axis. The optimal alignment is defined by the shortest warping path in a distance matrix. A warping path $W$ is a set of contiguous matrix indices defining a mapping between two time series. Even if there is an exponential number of possible warping paths, the optimal path is the one that minimizes the global warping cost. DTW can be computed using dynamic programming with time complexity $O(n^2)$ [298]. However, several lower bounding measures have been introduced to speed up the computation. [196] introduced the notion of upper and lower envelope that represents the maximum allowed warping. Using this technique, the complexity becomes $O(n)$. It is also possible to impose a *temporal constraint* on the size of the DTW warping window. It has been shown that these improve not only the speed but also the level of accuracy as it avoids the pathological matching introduced by extended warping

[299]. The two most frequently used global constraints are the Sakoe-Chiba Band and the Itakura Parallelogram. [312] introduced the FastDTW algorithm which makes a linear time computation of DTW possible by recursively projecting a warp path to a higher resolution and then refining it. A drawback of this algorithm is that it is approximate and therefore offer no guarantee to finding the optimal solution. In addition to dynamic warping, it may sometimes be useful to allow a global scaling of time series to achieve meaningful results, a technique known as *uniform scaling* (US). [131] proposed the scaled and warped matching (SWM) similarity measure that makes it possible to combine the benefits of DTW with those of US.

Other shape-based measures have been introduced such as the Spatial Assembling Distance (SpADe) [90]; it is a pattern-based similarity measure. This algorithm identifies matching *patterns* by allowing shifting and scaling on both temporal and amplitude axes, thus being scale robust. The DISSIM [129] distance has been introduced to handle similarity at various sampling rates. It is defined as an approximation of the integral of the Euclidean distance. One of the most interesting recent proposals is based on the concept of elastic matching of time series [216]. [217] presented an optimal subsequence matching (OSB) technique that is able to automatically determine the best subsequence and warping factor for distance computation; it includes a penalty when skipping elements. Optimality is achieved through a high computational cost; however, it can be reduced by limiting the skipping range.

*Edit-based*

Edit-based methods (also known as *Levenshtein distance*) has originally been applied to characterize the difference between two strings. The underlying idea is that the distance between strings may be represented by the minimum number of operations needed to transform one string into another, with insertion, deletion and substitution. The presence of outliers or noisy regions can thus be compensated by allowing gaps in matching two time series. [99] use the Longest Common Subsequence (LCSS) algorithm to tackle this problem. The LCSS distance uses a *threshold parameter* $\epsilon$ for point matching and a *warping threshold* $\delta$. A fast approximate algorithm to compute LCSS has been described in [51]. [364] normalized the LCSS similarity by the length of the time series and allowed linear transformations. [368] introduced lower-bounding measure and indexing techniques for LCSS. DTW requires the matched time series to be well aligned and its efficiency deteriorates with noisy data as, when matching all the points, it also matches the outliers distorting the true distance between sequences. LCSS has been shown to be more robust than DTW under noisy conditions [364]; this heavily depends on the threshold setting. [257] proposed the Fast Time Series Evaluation (FTSE) method for computing LCSS. On the basis of this algorithm, they proposed the Sequence Weighted Alignment model (Swale) that extends the $\epsilon$ threshold-based scoring techniques to include arbitrary match rewards and gap penalties. The Edit Distance on Real sequence (EDR) [86] is an adaptation of the edit distance to real-valued series. Contrary to LCSS, EDR assign penalties depending on the length of the gaps between the series. The Edit Distance with Real Penalty (ERP) [85] attempts to combine the merits of DTW and edit distance by using a *constant reference point*. For the same purpose, [243] submitted an interesting dynamic programming algorithm called Time Warp Edit Distance (TWED). TWED is slightly different from DTW, LCSS, or ERP algorithms. In particular, it highlights a parameter that controls a kind of stiffness of the elastic measure along the time axis. Another extension to the edit distance has been proposed in [259], it has been called the extended edit distance (EED). Following the observation that the edit distance penalizes all change operations with the same cost, it includes an additional term reflecting whether the operation implied characters that are more frequent, therefore closer in distance. A different approach for constraining the edit operations has been proposed in [91]; it is based on the Constraint Continuous Editing Distance (CCED) that adjusts the potential energy of each sequence to achieve optimal similarity. As CCED does not satisfy triangle inequality, a lower bounding distance is provided for efficient indexing.

*Feature-based*

These measures rely on the computation of a feature set reflecting various aspects of the series. Features can be selected by using coefficients from DFT [327] or DWT decompositions (cf. section 5.4.2) In [182], a likelihood ratio for DFT coefficients has been shown to outperform Euclidean distance. In [367], a combination of periodogram and autocorrelation functions allows to select the most important periods of a series. This can be extended to carrying out local correlation tracking as proposed in [269].

Concerning symbolic representations, [240] represent each symbol with a random vector and a symbolic sequence by the sum of the vectors weighted by the temporal distance of the symbols. In [124] weighted histograms of consecutive symbols are used as features. The similarity search based on Threshold Queries (TQuEST) [14] use a given threshold parameter $\tau$ in order to transform a time series into a sequence of *threshold-crossing* time intervals. It has however been shown to be highly specialized with mitigated results on classical datasets [106]. [28] proposed a Fourier-based approach, called WARP and making the using of the DFT phase possible, this being more accurate for a description of object boundaries.

An approach using ideas from shape and feature-based representations has been described in [247]. Typical local shapes are extracted with vector quantization and the time series are represented by histograms counting the occurrences of these shapes at several resolutions. Multiresolution Vector Quantized (MVQ) approximation keeps both local and global information about the original time series, so that defining a multi-resolution and hierarchical distance function is made possible.

*Structure-based*

Even if the previously cited approaches have been useful for short time series or subsequences applications, they often fail to produce meaningful results on longer series. This is mostly due to the fact that these distances are usually defined to find *local* similarities between patterns. However, when handling very long time series, it might be more profitable to find similarities on a more *global* scale. Structure-based distances [230] are thus designed to identify higher-level structures in series.

MODEL-BASED Model-based distances offer the additional advantage that prior knowledge about the generating process can be incorporated in the similarity measurement. The similarity can be measured by modeling one time series and determining the likelihood that one series was produced by the underlying model of another. Any type of parametric temporal model may be used. HMM with continuous output values or ARMA models are common choices [386]. However, best results are obtained if the model selected is related to the type of production that generated the data available. In [135], HMMs are combined with a piecewise linear representation. In [267] the distance between HMM is normalized to take into account the quality of fit of the series producing the model. [45] use the similarity-based paradigm where HMM is used to determine the similarity between each object and a pre-determinate set of other objects. For short time series, it is also possible to use regression models as proposed by [134].

Among other common choices for symbolic representations, we may cite MC [304], HMM with discrete output distributions [218], and grammar based models [10]. Alternatively to pairwise likelihood, the Kullback-Leibler divergence allows to have direct comparison of models [321].

COMPRESSION-BASED [201], inspired by results obtained in bioinformatics, defined a distance measure based on the Kolmogorov complexity called Compression-Based Dissimilarity Measure (CDM). The underlying idea is that concatenating and compressing similar series should produce higher compression ratios than when doing so with very different data. This approach appears to be particularly efficient for clustering; it has been applied to fetal heart rate tracings [98]. Following the same underlying ideas, [101] recently proposed a parsing-based similarity distance in order to distinguish healthy patients from hospitalized ones on the basis of various symbolic codings of ECG signals. By comparing the performances of several data classification methods, this distance is shown to be a good compromise between accuracy and computational efforts. Similar approaches have been undertaken earlier in bioinformatics [89] and several compression techniques – such as the Lempel-Ziv complexity [262] – have been successfully applied to compute similarity between biological sequences.

*Comparison of distance measures*

The choice of an adequate similarity measure highly depends on the nature of the data to analyze as well as application-specific properties that could be required. If the time series are relatively short and visual perception is a meaningful description, shape-based methods seems to be the appropriate choice. If the application is targeting a very specific dataset or any kind of prior knowledge about the data is available, model-based methods may provide a more meaningful abstraction. Feature-based methods seem more appropriate when periodicities is the central subject of interest and causality in the time series is not relevant. Finally, if the time series are long and little knowledge about the structure is available, the compression-based and more generally structure-based approaches have the advantage of being a more generic and parameter-free solution for the evaluation of similarity. Even with these general recommendations and comparisons for the selection of an appropriate distance measure, the accuracy of the similarity chosen still has to be evaluated. Ironically, it seems almost equally complex to find a good accuracy measure to evaluate the different similarities. However (cf. section 5.4.4), a crucial result when indexing is that any distance measure should lower bound the true distance between time series in order to preclude false dismissals [120]. Therefore the tightness of lower bound [194] appears to be the most appropriate option to evaluate the performance of distance measures as it is a completely hardware and implementation independent measure and offers a good prediction concerning the indexing performance. The accuracy of distance measures are usually evaluated within a 1-NN classifier framework. It has been shown by [106] that, despite all proposals regarding different kinds of robustness, the forty year old DTW usually performs better. Table 20 summarizes the properties of every distance measures reviewed in this chapter, based on our formalization of four types of robustness. It also determines whether the distance is a metric and indicates the computational cost and the number of parameters required.

| Distance measure | Scale | Warp | Noise | Outlier | Metric | Cost | P |
|---|---|---|---|---|---|---|---|
| **Shape-based** | | | | | | | |
| $L_p$ norms | | | | | ✓ | $O(n)$ | 0 |
| Dynamic Time Warping (DTW) | | ✓ | | | | $O(n^2)$ | 1 |
| LB_Keogh (DTW) | | ✓ | ✓ | | ✓ | $O(n)$ | 1 |
| Spatial Assembling (SpADe) | ✓ | ✓ | ✓ | | | $O(n^2)$ | 4 |
| Optimal Bijection (OSB) | | ✓ | ✓ | ✓ | | $O(n^2)$ | 2 |
| DISSIM | | ✓ | ✓ | | ✓ | $O(n^2)$ | 0 |
| **Edit-based** | | | | | | | |
| Levenshtein | | | | ✓ | ✓ | $O(n^2)$ | 0 |
| Weighted Levenshtein | | | | ✓ | ✓ | $O(n^2)$ | 3 |
| Edit with Real Penalty (ERP) | | ✓ | | ✓ | ✓ | $O(n^2)$ | 2 |
| Time Warp Edit Distance (TWED) | | ✓ | | ✓ | ✓ | $O(n^2)$ | 2 |
| Longest Common SubSeq (LCSS) | | ✓ | ✓ | ✓ | | $O(n)$ | 2 |
| Sequence Weighted Align (Swale) | | ✓ | ✓ | ✓ | | $O(n)$ | 3 |
| Edit Distance on Real (EDR) | | ✓ | ✓ | ✓ | ✓ | $O(n^2)$ | 2 |
| Extended Edit Distance (EED) | | ✓ | ✓ | ✓ | ✓ | $O(n^2)$ | 1 |
| Constraint Continuous Edit (CCED) | | ✓ | ✓ | ✓ | | $O(n)$ | 1 |
| **Feature-based** | | | | | | | |
| Likelihood | | | ✓ | ✓ | ✓ | $O(n)$ | 0 |
| Autocorrelation | | | ✓ | ✓ | ✓ | $O(n \log n)$ | 0 |
| Vector quantization | | ✓ | ✓ | ✓ | ✓ | $O(n^2)$ | 2 |
| Threshold Queries (TQuest) | | ✓ | ✓ | ✓ | | $O(n^2 \log n)$ | 1 |
| Random Vectors | | ✓ | ✓ | ✓ | | $O(n)$ | 1 |
| Histogram | | | ✓ | ✓ | ✓ | $O(n)$ | 0 |
| WARP | ✓ | ✓ | ✓ | | ✓ | $O(n^2)$ | 0 |
| **Structure-based** | | | | | | | |
| *Model-based* | | | | | | | |
| Markov Chain (MC) | | | ✓ | ✓ | | $O(n)$ | 0 |
| Hidden Markov Models (HMM) | ✓ | ✓ | ✓ | ✓ | | $O(n^2)$ | 1 |
| Auto-Regressive (ARMA) | | | ✓ | ✓ | | $O(n^2)$ | 2 |
| Kullback-Leibler | | | ✓ | ✓ | ✓ | $O(n)$ | 0 |
| *Compression-based* | | | | | | | |
| Compression Dissimilarity (CDM) | | ✓ | ✓ | ✓ | | $O(n)$ | 0 |
| Parsing-based | | ✓ | ✓ | ✓ | | $O(n)$ | 0 |

Table 1: Comparison of the distance measures surveyed in this chapter with the four properties of robustness. Each distance measure is thus distinguished as *scale* (amplitude), *warp* (time), *noise* or *outliers* robust. The next column shows whether the proposed distance is a metric. The cost is given as a simplified factor of computational complexity. The last column gives the minimum number of parameters setting required by the distance measure.

### 5.4.4 *Indexing*

An indexing scheme allows to have an efficient organization of data for quick retrieval in large databases. Most of the solutions presented involve a dimensionality reduction in order to index this representation using a spatial access method. Several studies suggest that the various representations differ but slightly in terms of indexing power [194]. However, wider differences arise concerning the quality of results and the speed of querying. There are two main issues when designing an indexing scheme: *completeness* (no false dismissals) and *soundness* (no false alarms). In an early paper, [120] list the properties required for indexing schemes:

1. It should be much faster than sequential scanning.
2. The method should require little space overhead.
3. The method should be able to handle queries of various lengths.
4. The method should allow insertions and deletions without rebuilding the index.
5. It should be correct, i.e. there should be no false dismissals.

As noted by [198] there are two additional desirable properties:

1. It should be possible to build the index within "reasonable time".
2. The index should be able to handle different distance measures.

A time series $X$ can be considered as a point in an $n$-dimensional space. This immediately suggests that time series could be indexed by Spatial Access Methods (SAMs). These allow to partition space into regions along a hierarchical structure for efficient retrieval. B-trees [29] on which most hierarchical indexing structures are based, were originally developed for one-dimensional data. They use prefix separators, thus no overlap for unique data objects is guaranteed. Multidimensional indexing structures – such as the R-tree [30] – use data organized in minimum bounding rectangles (MBR). However, when summarizing data in minimum bounding regions, the sequential nature of time series cannot be captured. Their main shortcoming is that wide MBR produce large overlap with a majority of empty space. Queries therefore intersect with many of these MBRs.
Typical time series contain over thousand datapoints and most SAM approaches are known to degrade quickly at dimensionality greater than 8-12 [79]. The degeneration with high dimensions caused by overlapping can result in having to access almost the entire dataset by random I/O. Therefore, any benefit gained when indexing is lost. As R-trees and their variants are victims of the phenomenon known as the *'dimensionality curse'* [50], a solution for their usage is to first perform dimensionality reduction. The X-tree (extended node tree), for example, uses a different split strategy to reduce overlap [34]. The A-tree (approximation tree) uses VA-file-style (vector approximation file) quantization of the data space to store both MBR and VBR (virtual bounding rectangle) lower and upper bounds [310]. The TV-tree (telescopic vector tree) is an extension of the R-tree. It uses minimum bounding regions (spheres, rectangles or diamonds, depending on the type of $L_p$ norm used) restricted to a subset of active dimensions. However, not all methods rely on SAM to provide efficient indexing. [272] proposed the use of suffix trees [156] to index time series. The idea is that distance computation relies on comparing prefixes first, so it is possible to store every series with identical

prefixes in the same nodes. The subtrees will therefore only contain the suffixes of the series. However, this approach seems hardly scalable for longer time series or more subtle notions of similarity. In [120] the authors introduced the GEneric Multimedia INdexIng method (GEMINI) which can apply any dimensionality reduction method to produce efficient indexing. [395] studied the problem of multi-modal similarity search in which users can choose between multiple similarity models depending on their needs. They introduced an indexing scheme for time series where the distance function can be any $\mathcal{L}_p$ norm. Only one index structure is needed for all $\mathcal{L}_p$ norms. To analyze the efficiency of indexing schemes, [166] considered the general problem of database indexing workloads (combinations of data sets and sets of potential queries). They defined a framework to measure the efficiency of an indexing scheme based on two characterizations: *storage redundancy* (how many times each item in the data set is stored) and *access overhead* (how many unnecessary blocks are retrieved for a query). For indexing purposes, envelope-style upper and lower bounds for DTW have been proposed [196]; the indexing procedure of short time series is efficient but similarity search typically entails more page reads. This framework has been extended [368] in order to index multidimensional time series with DTW as well as LCSS. [12] proposed the TS-tree, an indexing method offering efficient similarity search on time series. It avoids overlap and provides compact meta data information on the subtrees, thus reducing the search space. In [208], the use of an Incremental DFT Computation index (IDC-Index) has been proposed to handle streams based on a deferred update policy and an incremental computation of the DFT at different update speeds. However, the maintenance of the R*-tree for the whole streaming series might cause a constantly growing overhead and the latter could result in performance loss. It is also possible to use indexing methods to speed up DTW calculation; however, it induces a tradeoff between efficiency and I/O cost. However, [328] recently showed that for datasets that are large enough, the benefits of using DTW instead of Euclidean distance is almost null, as the larger the dataset, the higher the probability to find an exact match for any time series. They proposed an extension of the SAX representation – called indexable SAX (iSAX) – allowing to index time series with zero overlap at leaf nodes.

## 5.5 RESEARCH TRENDS AND ISSUES

Time series data mining has been an ever growing and stimulating field of study that has continuously raised challenges and research issues over the past decade. We discuss in the following open research issues and trends in time series data mining for the next decade.

STREAM ANALYSIS The last years of research in hardware and network research has witnessed an explosion of streaming technologies with the continuous advances of bandwidth capabilities. Streams are seen as continuously generated measurements which have to be processed in massive and fluctuating data rates. Analyzing and mining such data flows are computationally extreme tasks. Several papers review research issues for data streams mining [133] or management [139]. Algorithms designed for static datasets have usually not been sufficiently optimized to be capable of handling such continuous volumes of data. Many models have already been extended to control data streams, such as clustering [107], classification [173], segmentation [199] or anomaly detection [93]. Novel techniques will be required and they should be designed specifically to cope with the ever flowing data streams.

CONVERGENCE AND HYBRID APPROACHES A lot of new tasks can be derived through a relatively easy combination of the already existing tasks. For instance, [225] proposed three approaches, polynomial, DFT and probabilistic, to predict the unknown values that have not fed into the system and answer queries based on forecast data. This approach is a combination of prediction (cf. section 5.3.5) and query by content (cf. section 5.3.1) over data streams. This work shows that future research has to rely on the convergence of several tasks. This could potentially lead to powerful hybrid approaches.

EMBEDDED SYSTEMS AND RESOURCE-CONSTRAINED ENVIRONMENTS With the advances in hardware miniaturization, new requirements are imposed on analysis techniques and algorithms. Two main types of constraints should absolutely be met when hardware is inherently limited. First, embedded systems have a very limited memory space and can not have continuous access to it. However, most method use disk-resident data to analyze any incoming informations. Furthermore, sensor networks (which are frequently used in embedded systems) usually generate huge amounts of streaming data. So there is a vital need to design space efficient techniques, in terms of memory consumption as well as number of accesses. An interesting solution has been recently proposed in [393]. The algorithm is termed *autocannibalistic*, meaning that it is able to dynamically delete parts of itself to make room for new data. Second, as these resource-constrained environments are often required to be autonomous, minimizing energy consumption is another vital requirement. [43] has shown that sending measurements to a central site in order to process huge amounts of data is energy inefficient and lack scalability.

DATA MINING THEORY AND FORMALIZATION A formalization of data mining would drastically enhance potential reasoning on design and development of algorithms through the use of a solid mathematical foundation. [119] examined the possibility of a more general theory of data mining that could be as useful as relational algebra is for database theory. They studied the link between data mining and Kolmogorov complexity by showing their close relatedness. They conclude from the undecidability of the latter that data mining will never be automated, and therefore stating that "*data mining will always be an art*". However, a mathematical formalization could lead to global improvements of both reasoning and the evaluation of future research in this topic.

PARAMETER-FREE DATA MINING One of the major problems affecting time series systems is the large numbers of parameters induced by the method. The user is usually forced to "fine-tune" the settings in order to obtain best performances. However, this tuning highly depends on the dataset and parameters are not likely to be explicit. Thus, parameter-free systems is one of the key issue that has to be addressed. [201] proposed a first step in this direction by introducing a compression-based algorithm which does not require any parameter. As underlined by [119], this approach could lead to elegant solutions free from the parameter setting problem.

USER INTERACTION Time series data mining is starting to be highly dedicated to application specific systems. The ultimate goal of such methods is to mine for higher-order knowledge and propose a set of solutions to the user. It could therefore seem natural to include an user interaction scheme to allow for dynamic exploration and refinement of the solutions. An early proposal by [195] allows for relevance feedback in order to improve the querying process. From the best results of a query, the user is able to assign positive or negative influences to the series. A new query is then created by merging the series with respect to the user factors on which the system iterates. Few systems have tried to follow the same direction. However, an interactive mining environment allowing dynamic user exploration could increase the accessibility and usability of such systems.

exhaustive benchmarking A wide range of systems and algorithms has been proposed over the past few years. Individual proposals are usually submitted together with specific datasets and evaluation methods that prove the superiority of the new algorithm. As noted by [193], selecting those datasets may lead to *data bias* and showed that the performance of time series systems is highly data-dependent. The superiority of an algorithm should be tested with a whole range of datasets provided by various fields [106]. There is still a need for a common and exhaustive benchmarking system to perform objective testing. Another highly challenging task is to develop a procedure for real-time accuracy evaluation procedure. This could provide a measure of the accuracy achieved, thus allowing to interact with the system in real-time to improve its performance.

adaptive mining algorithm dynamics Users are not always interested in the results of a simple mining task and prefer to focus on evolution of these results in time. This actually represents the *dynamics* of a time series data mining system. This kind of study is of particular relevance in the context of data streams. [108] studied what are the distinctive features of analyzing streams are, rather than other kinds of data. They argued that one of the core issues is to mine *changes* in data streams. As they are of constantly evolving nature, a key aspect of the analysis of such data is to establish how an algorithm is able to adapt dynamically to such continuous changes. Furthermore, this could lead to ranking changes on the basis of relevance measures and contribute to the elaboration of methods to summarize and represent changes in the system. By finding a way to measure an approximate accuracy in real-time, it should be possible to imagine more "morphable" algorithms that could adapt dynamically to the nature of the data available on the basis of their own performances.

link to shape analysis Shape analysis has also been matter for discussion over the past few years. There is an astonishing resemblance between the tasks that have been examined; such as query by content [42], classification [191], clustering [228], segmentation [320] and even motif discovery [383]. As a matter of fact, there is a deeper connection between these two fields as recent work shows the numerous inherent link existing between these. [25] studied the problem of classifying ordered sequences of digital images. When focusing on a given pixel, it is possible to extract the time series representing the evolution of the information it contains. As this series is morphologically related to the series of the neighboring pixels, it is possible to perform a classification and segmentation based on this information. As presented above, [394] proposed to extract a time series from the contour of an image. They introduced the time series shapelets that represents the most informative part of an image and allows to easily discriminate between image classes. We can see from these works that both fields could benefit from each other. Even if only modest progress has been made in that direction, a convergence of both approaches could potentially lead to powerful systems.

# 6 MULTIOBJECTIVE OPTIMIZATION

Multiobjective approaches were designed to handle problems where several objectives are required to be optimized simultaneously. In order to achieve such a joint optimization, an alternative notion of optimality needs to be adopted. This concept was originally introduced by Francis Edgeworth in [113] and later generalized by the economist Vilfredo Pareto [270] and is called the *Pareto optimum*. As we will detail later, a Pareto solution is optimal in each direction of optimization. Therefore, multiobjective methods can be used to analyze and select between several potentially feasible options. Furthermore, the core strength of the multiobjective approach is that it allows a high degree of flexibility. Indeed, the optimal solution can perform extremely badly on a dimension as long as they perform extremely well on another. The distribution of such solutions is usually referred to as the *Pareto front* [83]. Multiobjective optimization has been applied to many real-world situations like natural resource management [248], medical diagnosis [31] and chemical engineering [44]. We refer the interested reader to recent reviews on multiobjective optimization and analysis techniques Ehrgott [115], Figueira et al. [123].

## 6.1 DEFINITIONS

Multiobjective minimization problems are defined by a given search space (sometimes called *decision space*) $S$ and a set of functions $F = \{f_1, ..., f_N\}$ to minimize over $S$. Formally, a multiobjective problem is defined by

$$\begin{cases} \min & F(x) = \{f_1(x), ..., f_N(x)\} \\ \text{s.t.} & x \in S \end{cases} \tag{6.1}$$

**Definition 19.** Let $S$ be a decision space and $F$ the set of functions of a multiobjective problem over $S$. The *criteria space* is defined as

$$C = \{(f_1(x), ..., f_N(x)) \mid x \in S\} \tag{6.2}$$

Usually, the ideal solution $x_{ideal}$, which is the global minimum for all criteria does not exist

$$\nexists x_{ideal} \in S,\ \forall n \in \{1, ..., N\},\ f_n(x_{ideal}) = \min_S f_n \tag{6.3}$$

Therefore, multiobjective problems cannot be solved with a single "perfect" solution, but rather with a *set* of *efficient solutions* called *Pareto solutions*.

### 6.1.1 *Pareto dominance*

As multiobjective optimization is based on finding a set of tradeoffs among potential solutions, we need a relaxed definition of similarity. An efficient solution is, therefore,

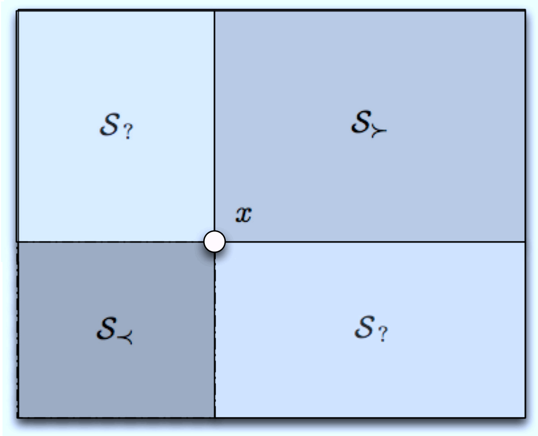



Figure 16: *Pareto dominance* relations for a minimization problem in a bi-criteria space. Any point $x$ of the criteria space divides it into three sub-spaces depending on the dominance relation. $\mathcal{S}_\prec$ contains the elements that dominates $x$ ($\forall y \in \mathcal{S}_\prec, y \prec x$). Elements of $\mathcal{S}_\succ$ are dominated by $x$ ($\forall y \in \mathcal{S}_\succ, x \prec y$). Finally, the elements of $\mathcal{S}_?$ simply cannot be compared to $x$ as they are not dominated nor dominate $x$ ($\forall y \in \mathcal{S}_?, x \nprec y \wedge y \nprec x$).

a solution that is not dominated in *every* objective. In other words, it is impossible to find another solution that jointly improves the complete set of criteria of an efficient solution.

**Definition 20.** Let $x$ and $y$ be two points of a search space $S$. We say that a solution $y$ is *Pareto dominated* by a solution $x$ (noted $x \preceq y$) if and only if it is dominated in every dimension. More formally

$$\forall n \in \{1, ..., N\},\ f_n(x) \leqslant f_n(y) \tag{6.4}$$

In that case $y$ is said to be a *weakly dominated* by $x$

**Definition 21.** We say that $x$ *strictly dominates* $y$ (noted $x \prec y$) iff

$$\begin{cases} \forall n \in \{1, ..., N\}, & f_n(x) \leqslant f_n(y) \\ \exists n_0 \in \{1, ..., N\}, & f_{n_0}(x) < f_{n_0}(y) \end{cases} \tag{6.5}$$

We say that a solution $y$ is *strongly dominated* by a solution $x$ (noted $x \prec\prec y$) if and only if it is strictly dominated in every dimension. More formally

$$\forall n \in \{1, ..., N\},\ f_n(x) < f_n(y) \tag{6.6}$$

The dominance relation $\prec$ induces only a partial order on the criteria space, as shown in Figure 16. For any element $x$, the criteria space is divided into three regions depending on the dominance relation between $x$ and the corresponding subspaces. We call these three subspaces $\mathcal{S}_\prec$, $\mathcal{S}_\succ$ and $\mathcal{S}_?$. $\mathcal{S}_\prec$ contains the elements that dominate $x$ ($\forall y \in \mathcal{S}_\prec, y \prec x$). $\mathcal{S}_\succ$ is the subspace whose elements are dominated by $x$ ($\forall y \in \mathcal{S}_\succ, x \prec y$). Finally, the elements of $\mathcal{S}_?$ just cannot be compared to $x$ as they are not dominated nor dominate $x$ ($\forall y \in \mathcal{S}_?, x \nprec y \wedge y \nprec x$).

**Definition 22.** The set of non-dominated (or efficient) elements of $S$ is called the *Pareto front*.

Solving a multiobjective problem can thus be summarized as the discovery of the Pareto front. Figure 17 depicts a search space in the bi-objective case. We can clearly see in this figure the Pareto front which emerges from all the non-dominated solutions. From a strict point of view, none of the Pareto solutions can be preferred to others. As we can see, this approach provides a high degree of flexibility, as all directions of optimization are allowed.

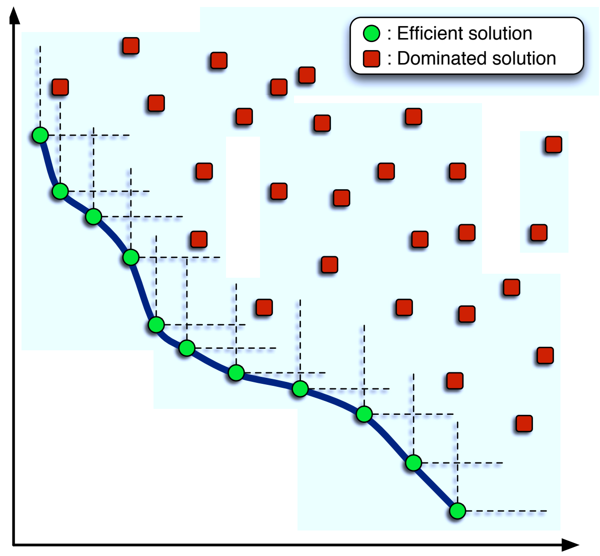



Figure 17: Efficient solutions and dominated solutions for a bi-criteria minimization problem. We can clearly see the *Pareto front* of non-dominated solutions. The dotted lines define the sub-spaces which are dominated by these solutions.

### 6.1.2 *Chebyshev norms*

Now, if we look at the problem from the opposite side, each non-dominated element of the Pareto front is the best solution of a mono-objective problem with respect to a set of weights.

**Definition 23.** Let $\Lambda$ be the subset of $[0;1]^N$ such that

$$\forall \lambda = (\lambda_1, \ldots, \lambda_N) \in \Lambda, \ \sum_i \lambda_i = 1 \tag{6.7}$$

The norm defined over a criteria space $C$ by

$$\forall x \in C, \|x\|_\lambda = \max_i \lambda_i x_i \tag{6.8}$$

is called the *weighted Chebyshev norm* given the set of weights $\lambda \in \Lambda$.

**Proposition 24.** *A solution $x$ belongs to the Pareto front if and only if there exists a set of weights $\lambda \in \Lambda$ such that*

$$x = \operatorname*{argmin}_{y \in C} \|y\|_\lambda \tag{6.9}$$

Figure 18 illustrates this duality between a multiobjective problem and a set of mono-objective problems. The efficient configurations $x_a$, $x_b$ and $x_c$, are the best solutions of an associated monoobjective minimization problem given respectively by norms $\mathcal{N}_a$, $\mathcal{N}_b$ and $\mathcal{N}_c$. For each of them, the associated sets of weights defines their weighted Chebyshev norm and, therefore, different directions of optimization.

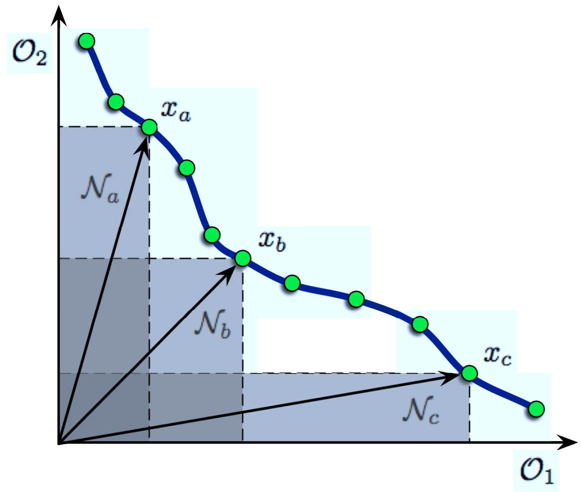



Figure 18: Three efficient solutions $x_a$, $x_b$, $x_c$ and their corresponding induced weighted Chebyshev norms $\mathcal{N}_a, \mathcal{N}_b, \mathcal{N}_c$ in a bi-objective problem. Each point of the Pareto front is thus the best solution of a mono-objective problem weighted by its corresponding Chebyshev norm.

## 6.2 algorithmic approaches

Over the years, various methods and algorithms have been proposed to solve multi-objective optimization problems. Overall, these methods can be divided in two broad categories depending on their approach. First, the *complete methods* allow to find the best optimal solutions (global minimum), but do not provide any limit on their execution time. Therefore, they sometimes can not be feasible depending on the complexity of the problem at hand. Second, *meta-heuristics* allow to obtain approximate solutions in a fixed amount of time. However, the outcome of such algorithms is usually not deterministic and can not provide absolute guarantees on the quality of found solutions. As usually the case with combinatorial problems, we can see that there is a tradeoff between the *efficiency* (execution time) and the *effectiveness* (quality of solutions) of the algorithms.

When the properties of the problem are of limited complexity (small number of objectives, linear objective functions, restricted criteria space), complete methods can provide an optimal solution to multiobjective optimization. In these cases, classical optimization tools such as dynamic programmation, the $A^*$ search algorithm or *branch and bound* algorithms are preferred because they guarantee the optimality of solutions as they span the search space entirely. However, for problems that require more than two objectives or with a wide search space, such methods are impossible to use because of the search complexity and combinatorial explosion.

The *metaheuristics* allow to provide a workaround for these shortcomings by providing approximate solutions in a reasonable amount of time, even if the solutions are usually suboptimal. The effectiveness of such methods can not be proved theoretically but are exhibited through empirical experiments. The proposed metaheuristics for solving multiobjective optimization problems broadly falls into two distinct categories. First, the *neighborhood* methods try to refine iteratively a single configuration in order to converge towards the Pareto front. Second, the *population* methods use a set of configurations that interact together in order to optimize jointly the overall quality of the set. This approach has been the most studied in the literature, as exemplified by the popularity of *genetic algorithms* (GA) in multiobjective optimization. We redirect the interested readers to reviews on the proposed approaches for multiobjective combinatorial optimization problems Ehrgott and Gandibleux [116].

## 6.3 APPLICATIONS

Multiobjective optimization has been applied in a wide variety of scientific fields, often towards a system that could help in decision-making processes. Multicriteria methods can thus be used to analyze and select between different potentially feasible water resources development options [3] where it has been applied to many real-world situations [143] based on criteria such as environmental protection, water demand, regional cooperation. Similar lines of research have been followed for conceptual runoff [322] or hydrologic model calibration [392], for fisheries management [242] or environmental decision making [204]. Multicriteria analysis has also been widely used for agricultural resource management [161], either for selecting the optimal multiattribute alternative or for solving multiobjective planning problems. Romero and Rehman [308] underlined the suitability of MCDM techniques to natural resource management. They further presented the theoretical aspects of multicriteria techniques and detailed practical considerations for the management of agricultural systems [303]. A good review of multicriteria decision analysis applied to forest management and other natural resources has been presented in [248].

A lot of research has been devoted to applying multiobjective techniques to assist medical diagnosis. Belacel developed a fuzzy multicriteria classification method called PROAFTN [31] which is a supervised classification scheme. This technique is used to solve the *nominal sorting problematic* [278] where unordered categories are represented by reference objects called *prototypes*. They also showed its applicability to aid diagnosis of strocytic and bladder tumours [32] as well as acute leukemia [33]. Based on a patient's symptoms, Technique Ordered Preference by Similarity to the Ideal Solution (TOPSIS) is one of the widely used methods in medical diagnosis systems [295]. More recently, Zhang et al. [407] developed a linear programming approach for improving the accuracy of medical diagnosis as well as prognosis. Application of multiobjective optimization to chemical engineering has also seen a flourishing literature in the past years [44], with problems related to optimization of chemical processes design **[323]**, polymerization reactions [357], waste treatment [347] and air polution control [300]. Multiobjective optimization has even been applied to beer dialysis **[400]** in order to preserve a beer's unique taste even with lowered alcohol content.

# Part III

# MULTIOBJECTIVE TIME SERIES (MOTS) MATCHING

# 7 MOTS FRAMEWORK

As we discussed in Section 2.4, studying the properties of musical atoms and auditory perception has taught us that we process audio features in both a temporal and multidimensional manner. As we also discussed in Section 3.3, we are able to perceive uncorrelated features simultaneously and perform flexible similarity assessments. Hence, it seems that a single measure would be unlikely to convey such perceptual similarity (as already pointed out by several authors [111, 110, 363]). Therefore, we believe that both this multidimensional nature and the ability to perceive complex temporal structures shall be taken into account in similarity matching. However, no current retrieval system seems to address these limitations (even outside the realm of audio matching). For example, audio retrieval techniques usually just borrow from other fields such as pattern recognition in order to obtain a single audio similarity measure. While they clearly address many engineering problems, they do not expressly address the multidimensional issues involved in the similarity of timbre. Hence, we wish to incorporate time series information and at the same time seek a multidimensional assessment of similarity. Motivated by these observations, we introduce the generic *MultiObjective Time Series* (MOTS) matching problem. This problem can be applied to any problem in which various time series should be matched jointly, without favoring any dimension in the process. The goal of MOTS matching is therefore to provide a flexible comparison of multiple time series.

Equipped with the basic notions of time series matching and multiobjective optimization, we begin by introducing the general MOTS matching problem (Section 7.1). We indicate the core differences between this novel problem and multivariate matching (Section 7.2) and briefly discuss its complexity. We then introduce two algorithms to solve this problem (Section 7.3) and show their efficiency on massive sets of data (Section 7.4). We compare the efficiency of these algorithms between real and synthetic sets of data (Section 7.4.2). Finally, we discuss the application of the MOTS framework to audio retrieval settings. We show that, based on this framework, we can easily construct two innovative audio querying paradigms (Section 7.5), that allow to go beyond the traditional audio query applications.

## 7.1 PROBLEM DEFINITION

**Problem 25.** A *Multiobjective Time Series (MOTS)* matching problem is defined as finding the efficient elements of a database that jointly minimize a set of time series distances

$$\begin{cases} \min \mathcal{D}_Q^k(\mathcal{S}) & k \in \{1, \dots, K\} \\ \text{s.t. } \mathcal{S} \in DB \end{cases} \tag{7.1}$$

with $Q$ the query represented by a set of $K$ time series and $\mathcal{S}$ the elements of the database $DB$ which contains time series corresponding to the same objectives as the query. Finally, $\mathcal{D}_Q^k(\mathcal{S})$ is the similarity between the $k^{th}$ feature represented by time series $Q_k$ and $\mathcal{S}_k$, i.e. $\mathcal{D}_Q^k(\mathcal{S}) = \mathcal{D}(Q_k, \mathcal{S}_k)$ (cf. Definition 5).

It is necessary to note here that this problem is not a problem of *optimization* (as are usual multi-objective methods). Indeed, the elements in the database are fixed and, therefore, there is no *feature space*. However, we can already see now that part of the computational complexity of this problem arises from objective functions $\mathcal{D}_Q^k(\mathcal{S})$ which represent time series distances. As we discussed in Section 5.4.3, time series similarity is a remarkably subtle concept that can entail a high computational complexity. Furthermore, because of the multiobjective nature of this problem, it is impossible to gain straightforward efficiency from traditionnal time series indexing methods. Indeed, these techniques provide most of their pruning power by avoiding computation of irrelevant parts of the search space.

As we will discuss further in Section 7.2, it is noteworthy here to understand the fundamental differences between multivariate time series matching (extensively studied in literature) and our multiobjective problem. Multivariate problems usually imply that the series are somehow statistically linked and attempt to find a single similarity measure to compare a *set* of time series. This allows to circumvent the problem of pruning power raised by the notion of Pareto dominance, which is the second aspect of computational complexity in the MOTS problem. Indeed, multiobjective problems allow the optimization of objectives that can be conflictive with each other. Finding the most similar item $\mathcal{S}^*$ to a MOTS query requires to jointly minimizing the distances between two sets of time series.

$$\mathcal{S}^* = \underset{\mathcal{S}}{\operatorname{argmin}} \left\{ \left( D_Q^k(\mathcal{S}) \right), k = 1, ..., K \right\} \tag{7.2}$$

As the *ideal point* $\mathcal{S}^*$ which simultaneously optimizes all criteria usually does not exist, solving this problem turns out to finding the set of *tradeoff solutions* that offer different compromises among objectives. A solution $\mathcal{S}$ is optimal if there is no other solution in the search space that achieves similarities higher than $\mathcal{S}$ on *every* criterion $\mathcal{D}_Q^k(\mathcal{S})$. Therefore, we can not just rule out a portion of the database if it performs poorly on one objective, as it could contain the best element in another objective. This implies that if we want to know which elements belong to the exact Pareto front, we should evaluate the complete set of distances for every objective, thus degenerating to *brute force* analysis. Figure 19 illustrates these concepts. The query is a set of time series input to the system. The query is at the origin of the criteria space as distances with itself are null in every objective. There is no element in the database that perfectly matches these two time series. Solution $\mathcal{A}$ is the best match for objective $\mathcal{O}_1$. As we can see, its first time series is closely similar to that of the query. Solution $\mathcal{B}$ is respectively the best match for objective $\mathcal{O}_2$. Finally, element $\mathcal{C}$ is the best solution for the associated mono-objective problem with equal weights. We can see that it is not closely similar to any objective, which exhibits the relevance of our approach. Indeed, the MOTS matching allows joint queries on several dimensions without favoring any of them during the search. Therefore, this approach is an appropriate model when the relative weights of each objective cannot be known in advance, which is particularly relevant for audio perception (cf. Section 5). Depending on the problem, regions of the Pareto front might be preferred to others, according to personal preferences as we will discuss in Section 7.7. Finally, it is already appealing to note here that the objective functions $\mathcal{D}_Q^k(\mathcal{S})$ can be defined differently depending on the feature being studied. Therefore, the distance function in each dimension can be tailored to fit its corresponding feature. It is even possible to assign multiple objectives for the same time series features, with

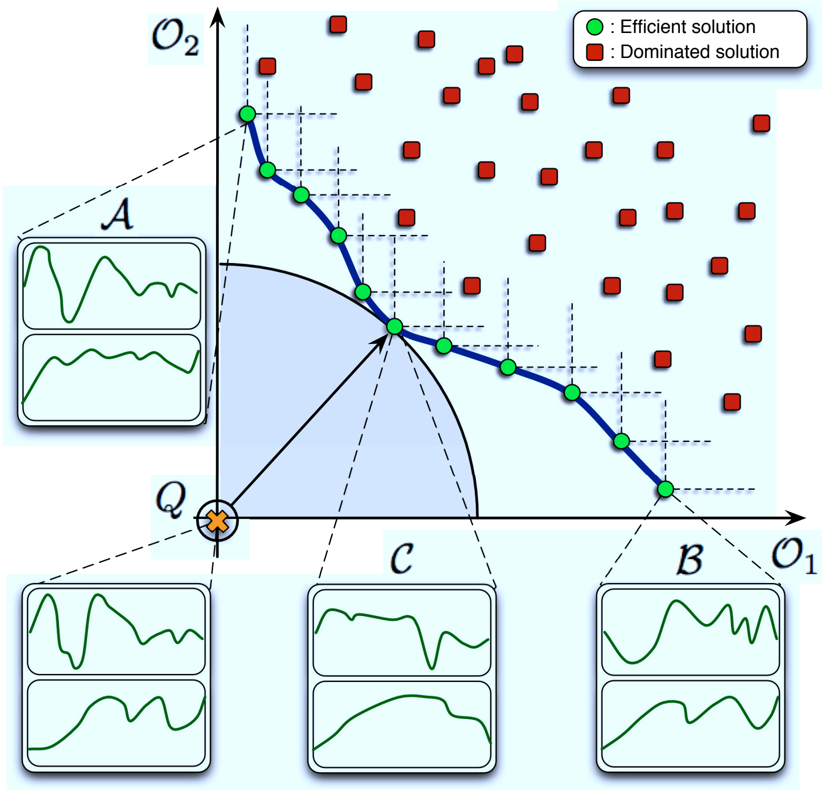



Figure 19: Illustration of the MOTS matching problem in a bi-objective context. The query Q is at the origin of the space and is represented by a set of time series that have to be matched jointly. Solution $\mathcal{A}$ is the best match for objective $\mathcal{O}_1$, as we can see the first time series is closely similar to that of the query. Solution $\mathcal{B}$ is respectively the best match for objective $\mathcal{O}_2$. The element $\mathcal{C}$ would be the best solution for the weighted monobjective problem given a set of equal weights. We can see that it is not closely similar to any objective, which motivates the use of multiobjective optimization.

each dimension corresponding to a different measure of similarity over the same feature.

## 7.2 COMPARISON TO MULTIVARIATE MATCHING

We try to demonstrate the fundamental differences between a multivariate matching algorithm (extensively studied in the literature) and our multiobjective problem. First, as we said earlier, multivariate analysis generally relies on the premise that the set of time series is somehow statistically linked. Sometimes, inference is even possible between the different series of the set. In this manner, multivariate analysis provides mechanisms to obtain trends *across* multiple dimensions and take into account the effect of all variables observed. Oppositely, multiobjective matching allows the optimization of objectives that can be uncorrelated and even conflictive with each other. As we have seen in Section 3.3, we can *perceive* conflictive temporal evolutions, which makes multiobjective approaches even more appealing. Furthermore, multivariate matching usually tries to find a single measure that could encapsulate the similarity on the *whole set* of time series. In this line of thought, a multivariate search can be seen as equivalent to a mono-objective problem spanning several dimensions. Indeed, the multivariate case can be seen as a reduction to a weighted multiobjective search. Hence, weighting and merging the objectives could allow to circumvent the problem of pruning power raised by the notion of Pareto dominance. Finally, the number of dimensions of a multivariate problem is usually fixed whereas the multiobjective approach can work on different subsets of objectives. We illustrate these concepts by trying to find the solution of the similarity problem exposed in Figure 5 with a multivariate nearest-neighbor approach. We can see in Figure 20 that the computation of the Euclidean distance ($\mathcal{L}_2$) on each feature gives a slight difference in results (because of the small oscillating segment in *loudness*). If we try to find which elements are similar to $\mathcal{S}_2^1$, the system will rate element $\mathcal{S}_1^1$ as being the most similar and then element $\mathcal{S}_2^2$ as being less similar. Therefore, this approach impose an implicit preference towards the *pitch* of different sounds in similarity matching.

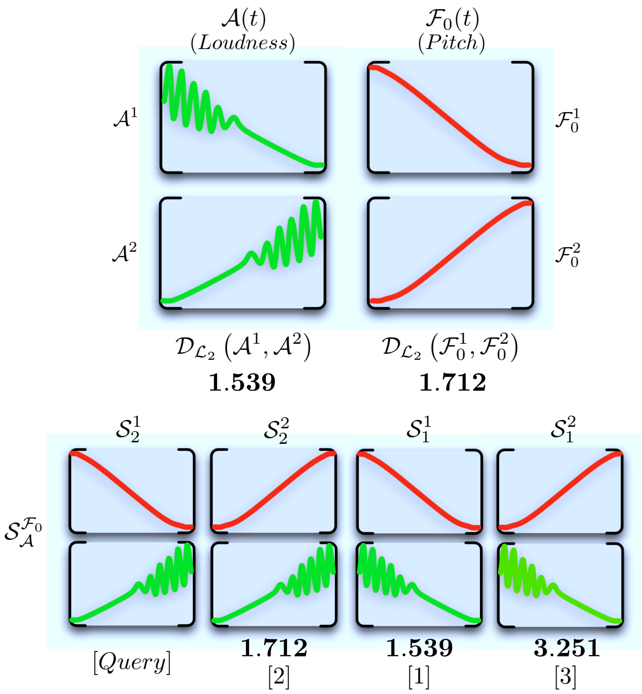



Figure 20: Trying to find the solution to the similarity problem exposed in Figure 5 with a multivariate nearest-neighbor approach. The system will order element $\mathcal{S}_1^1$ as being the most similar to $\mathcal{S}_2^1$, as the distance is slightly different between the two features, Therefore, there is an implicit preference towards the *pitch* of different sounds in similarity matching.

Now we try to solve the same similarity problem with a MOTS approach in Figure 21. This time, if we try to find which elements are more similar to $\mathcal{S}_2^1$, the system will isolate the problem depending on its two underlying dimensions. Therefore, element $\mathcal{S}_1^1$ and $\mathcal{S}_2^2$ are selected as *efficient* as they are not dominated in any dimension. However, there is no ranking between these two elements and they are treated as being "equally efficient" to the similarity towards $\mathcal{S}_2^1$. Only element $\mathcal{S}_1^2$ is clearly exhibited as being least similar, as it is dominated by the others. Therefore, there is no imposed preference towards any dimension of different sounds in similarity matching. The MOTS matching treats the two dimensions separately and equally.

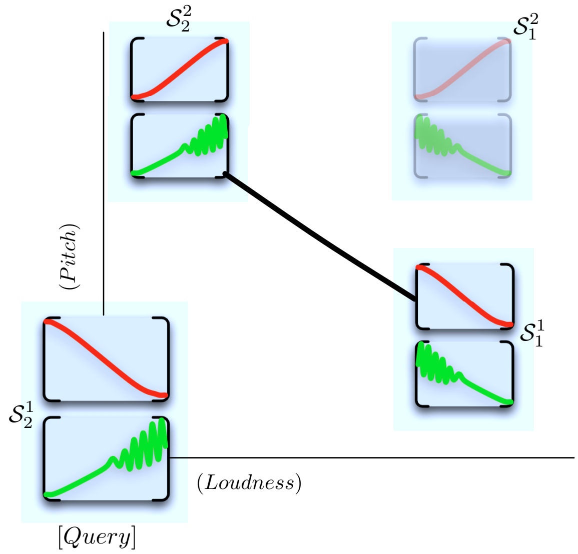



Figure 21: Trying to find the solution of the similarity problem exposed in Figure 5 with a MOTS approach. If we seek to find which elements are more similar to $\mathcal{S}_2^1$, the system will divide the problem in its two inherent dimensions. Therefore, element $\mathcal{S}_1^1$ and $\mathcal{S}_2^2$ are selected as being the most similar as they are not dominated. Only element $\mathcal{S}_1^2$ is exhibited as being least similar. Therefore, there is no implicit preference towards any dimension in similarity matching.

## 7.3 ALGORITHMS

Because of the ever-growing size of storage capacities, linear scan of an entire database has become unacceptable. Hence, it would be highly desirable to obtain a search method with sublinear time complexity. We introduce two algorithms that can handle the MOTS matching problem. However, because of the novelty of this approach, no competing method exists to evaluate the efficiency of our algorithms. Therefore, the *multiobjective brute force* algorithm will be our testing baseline. This approach requires to compute every distance in each objective and then extract the Pareto front from the full distance matrix. We describe this reference procedure in Algorithm 7.1.

**Algorithm 7.1** Brute force multiobjective time series matching algorithm

```
multiobjectiveBruteForce(Q, db)
  for i ∈ [1...size(db)]
    for k ∈ [1...N_obj]
      compute D_Q^k(S_i)
    end
  end
  P ← extractParetoFront(D_Q(S_j))
end
```

### 7.3.1 *Multiobjective early abandon*

As we discussed in Section 7.1, the complexity of the MOTS problem essentially lies in the repeated computations of time series distances. A natural idea would be, therefore, to find a way to restrict the amount of distance computations. Instead of computing the distances for every series and each objective, we would like to drop calculations as soon as we are confident that the corresponding element is dominated. This technique is known as *early abandon*. However, we have to make fundamental modifications in order to account for the multiobjective nature of our problem. Indeed, early abandon in a mono-objective setting is based on comparing the current similarity against the best distance known so far. However, in a multidimensional context where we seek a set of efficient solutions, we cannot simply compare the current distances to a single reference. Therefore, a first turnaround would be to maintain a current working Pareto front with which to compare the successive elements. However, this approach would require to perform several verifications of Pareto dominance at each step of a distance computation. Unfortunately, this verification is a computationally intensive operation. Therefore, it would be preferable to obtain an *approximate distance* for every elements beforehand. That way, we could perform the Pareto verification on these approximations and only compute the complete distances of potentially efficient solutions. The approximate distance should be *lower-bounding*, ie. it should underestimate the true distance.

$$\mathcal{D}^{k}_{approx}\left(\mathcal{S}_i\right) \leqslant \mathcal{D}^{k}_{true}\left(\mathcal{S}_i\right)\ \forall k \in \left[1, \ldots N_{obj}\right] \tag{7.3}$$

With this property, we can prune elements as we are sure that they can only perform *worse* than their current position. In simpler words, if a set of lower-bounding distances is dominated, then we are sure that the corresponding set of true distances is dominated. Therefore, we need to obtain *simplified representations* for the collection of time series that can provide a more efficient distance computation. If these representations are coarse enough, they can account for several time series at the same time. In order to obtain such properties, we can use the SAX representation Lin et al. [233] that performs a temporal and amplitude quantification of the series. In this model, the series are first divided into a set of equal-sized temporal steps. Then, the average of the time points contained in each step $i$ is computed and matched to an alphabet of reduced size.

$$\bar{\mathcal{T}}_i = \alpha\left(\frac{w}{n}\sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i}\mathcal{T}_j\right) \tag{7.4}$$

with $n$ the length of the original series, $w$ the number of resulting temporal steps ($w \ll n$) and $\alpha(x)$ a function that matches $x \in \mathbb{R}$ to a discrete alphabet (amplitude quantification). Based on this representation, the iSAX index [328] provides an efficient tree-like index for time series. The idea behind this index is that each level of the tree provides a finer representation of the series, by increasing the size of the amplitude alphabet. Figure 22 illustrates this construction. The series are dived into 8 equal-sized temporal steps. At the first level, the series are quantified by using an alphabet of two elements $\{0, 1\}$. Then, at the subsequent levels, the series are refined by using a larger alphabet $\{00, 01, 10, 11\}$. Obviously, each node in the tree accounts for a whole set time series from the database. Hence, if we take the first-level of this representation, we obtain a set of *prototypical bins* of reduced cardinality for the complete database.

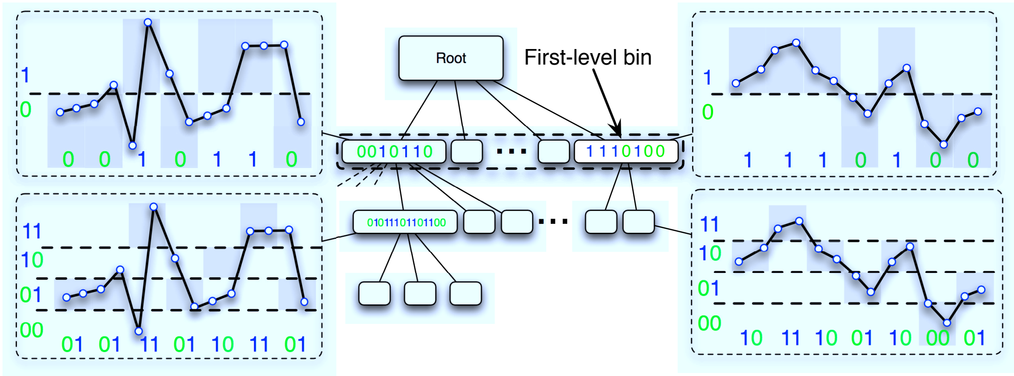



Figure 22: Construction of the quantified bins for time series and computation of the 1st-level distance for a query time series.

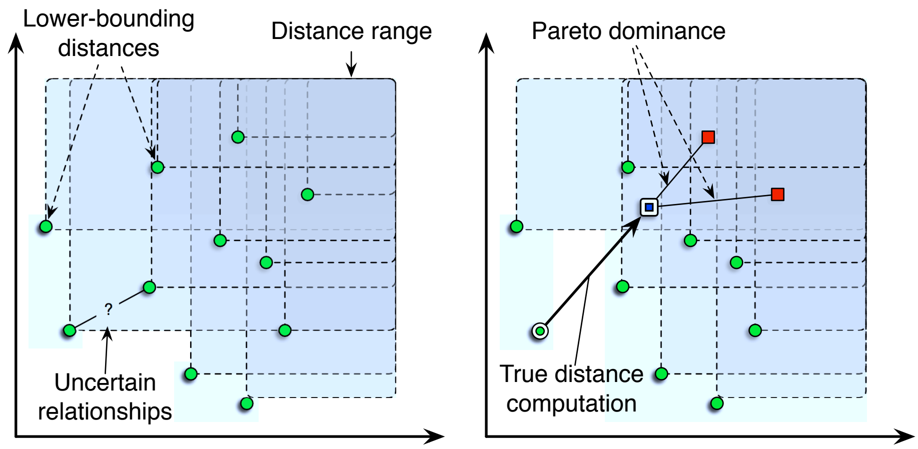



Figure 23: The approximate lower bounding distances in the criteria space and a set of relationships that can or can not be computed

Then, the lower-bounding distance between a query Q and a bin representation $\bar{\mathcal{B}}_x$ can be obtained by first transforming the query into the same representation $\bar{Q}$ and then computing

$$\mathcal{D}_{\mathrm{approx}}\left(\bar{Q}, \bar{\mathcal{B}}_x\right) = \sqrt{\frac{n}{w}}\sqrt{\sum_{i=1}^{w}\left(\mathcal{D}\left(q_i, b_i\right)\right)^2} \tag{7.5}$$

Hence, with this construction, we can obtain the lower bounding position of every element in the database, as illustrated in Figure 23. At first glance, it would seem tempting to use these approximate distances to perform a direct assessment of Pareto efficiency. However, it is important to understand that these distances are just lower-bound *approximations*. Therefore, the true dominance relations are still uncertain. This is exhibited in Figure 23 with an outlined relationship. It turns out that the final distance of the potentially dominating element is much higher. However, when its true distances are computed, we are sure that it dominates some of the approximate positions.

The final implementation is presented in Algorithm 7.2. We start by transforming the query into the quantified representation. Then, we compute the first level distances

**Algorithm 7.2** MOTS matching algorithm with early abandon

```
multiobjectiveEarlyAbandon(Q, db, idx)
  // Quantify the query
  Q̄^{k∈[1...N_obj]} = { (w/n) Σ_{j=(n/w)(i−1)+1}^{(n/w)i} Q_j^k, i ∈ [1...w] }
  // Compute query-to-bin distances
  for k ∈ [1...N_obj]
    for b ∈ [1...N_bins^k]
      aDist^k_{i∈B_b^k} = D_approx(Q̄^k, B̄_b^k)
    end
  end
  P = ∅
  // Perform multiobjective abandon
  for i ∈ [1...size(db)]
    for k ∈ [1...N_obj]
      if isDominated(aDist_i, P)
        abandon;
      else
         aDist_i^k = D_Q^k(S_i)
    end
    add(S_i, P);
    P = extractParetoFront(P);
  end
```

for all bins. We store these distances for corresponding elements in the distance matrix $aDist$. We then create an empty Pareto front $\mathcal{P}$ and iterate over the elements of the database. When evaluating an element, as soon as it is dominated by the current Pareto front, we abandon computations. If all the distances have been computed, then the current element is potentially efficient. Therefore, we add this element to the current Pareto front. We compute the new front by removing the eventual dominated points (as the newly added item might dominate existing solutions in the current front). When all the elements of the database have been verified, $\mathcal{P}$ contains the final Pareto front.

### 7.3.2 *Hyperplane search*

One of the main problem of the previous algorithm is that it still requires frequent verifications of the Pareto optimality. Hence, it would be wiser to find a less expensive theoretical limit to drop computations of the distance measures. Therefore, our main idea is to construct an approximate Pareto hyperplane $\mathcal{P}$ to act as our theoretical limit. We can obtain this hyperplane by using 1-NN queries from efficient time series indexing for each objective (such as the iSAX index presented in the previous section). These queries will give us boundary elements of the final Pareto front. This is straightforward from the fact that these elements cannot be dominated as they have the smallest distance in one of the objectives.

$$\forall \mathcal{S}_i, \exists k \mid \forall \mathcal{S}_j, \mathcal{D}_Q^k(\mathcal{S}_i) < \mathcal{D}_Q^k(\mathcal{S}_j) \Rightarrow \mathcal{S}_i \in \mathcal{P} \tag{7.6}$$

Hence, we can prune elements whose approximate distances are dominated by this hyperplane. This can be computed straightforwardly if we obtain the hyperplane normal. We show how to compute this normal efficiently by avoiding an expensive least-squares minimization. The normal of a hyperplane can be defined in the following manner

**Proposition 26.** *Given a nonzero vector* $\mathbf{n}$ *in* $\mathbb{R}^m$ *and a point* $\mathbf{p} \in \mathbb{R}^m$*, the hyperplane perpendicular to* $\mathbf{n}$ *through* $\mathbf{p}$ *is the set of all* $\mathbf{x} \in \mathbb{R}^m$ *such that*

$$(\mathbf{x} - \mathbf{p}) \cdot \mathbf{n} = 0 \tag{7.7}$$

Therefore, if we want to find the normal of hyperplane $\mathcal{H}$, we must find the vector $\mathbf{n}_p \in \mathbb{R}^m$ satisfying

$$\mathcal{P}\mathbf{n}_p = \mathbf{0_m} \tag{7.8}$$

where $\mathcal{P}$ is a $k \times m$ matrix and $\mathbf{0_m}$ is a $m \times 1$ zero vector. $\mathcal{P} = [p_1, \ldots, p_k]$ is the set of Pareto points defining the hyperplane $\mathcal{H}$ (in our case $p_i$ will be the 1-NN result for the $i^{th}$ objective). In order to obtain this vector, we must solve

$$\mathbf{n}_p = \underset{\mathbf{v}}{\operatorname{argmin}} \left( \mathbf{v}^T \mathcal{P}^T \mathcal{P} \mathbf{v} \right) \tag{7.9}$$

Alone, this equation yields the trivial solution $\mathbf{n}_p = \mathbf{0_m}$ which we obviously want to avoid. To avoid this case, we can add the constraint $\|\mathbf{n}_p\| = 1$, which can be rewritten as $1 - \mathbf{n}_p^T \mathbf{n}_p = 0$. Therefore, in order to find the best value for $\mathbf{n}_p$, we can use the Lagrange multipliers and solve

$$\frac{\delta}{\delta \mathbf{n}_p} \left( \mathbf{n}_p^T \mathcal{P}^T \mathcal{P} \mathbf{n}_p + \lambda \left( 1 - \mathbf{n}_p^T \mathbf{n}_p \right) \right) = 0 \tag{7.10}$$

After applying the derivation, we obtain the characteristic equation $\left(\mathcal{P}^T \mathcal{P} - \lambda E\right) \mathbf{n}_p = 0$. Therefore, we know that $\mathbf{n}_p$ is an eigenvector of $\left(\mathcal{P}^T \mathcal{P}\right)$ and $\lambda$ is an eigenvalue. However, we can not control the orientation of the normal (as any hyperplane possess two oppositely oriented normal vectors). Furthermore, this also requires to compute some eigenvectors with potentially large dimensionality which can be expensive. In order to alleviate both problems at the same time, we have to slightly modify the original constraint. For that purpose, we introduce a direction vector $\mathbf{d}$ that will ensure the orientation of the normal vector. Therefore, as we constrain the normal vector $\mathbf{n}_p$ to have the same orientation as $\mathbf{d}$. We can write this constraint as $\left(1 - \mathbf{d}^T \mathbf{n}_p\right)^2 = 0$. Hence, we must now solve

$$\mathbf{n}_p = \underset{\mathbf{v}}{\operatorname{argmin}} \left( \mathbf{v}^T \mathcal{P}^T \mathcal{P} \mathbf{v} + \left( 1 - \mathbf{d}^T \mathbf{v} \right)^2 \right) \tag{7.11}$$

By using the same reasoning than previously, we can find the extreme value by solving

$$\frac{\delta}{\delta \mathbf{n}_p} \left( \mathbf{n}_p^T \mathcal{P}^T \mathcal{P} \mathbf{n}_p + \left( 1 - \mathbf{d}^T \mathbf{n}_p \right)^2 \right) = 0 \tag{7.12}$$

Therefore, by taking the same matrix derivatives and simplifying, we obtain the normal by computing

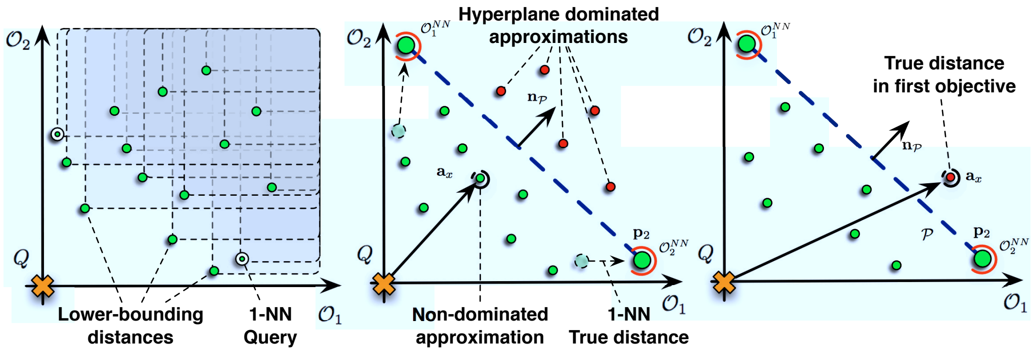



Figure 24: Geometric interpretation of the *multiobjective hyperplane search* algorithm

$$\mathbf{n}_p = \left([\mathcal{P},\mathbf{d}]\,[\mathcal{P},\mathbf{d}]^T\right)^{-1}\mathbf{d} \tag{7.13}$$

where $[\mathcal{P},\mathbf{d}]$ is the matrix obtained by concatenating matrix $\mathcal{P}$. In our implementation, we use $\mathbf{d} = \max_j(p_i^j)$, $p_i \in \mathcal{P}$ to ensure the orientation of the resulting normal. The distance of any point $a_x$ relative to the approximate Pareto front $\mathcal{P}$ is then defined as

**Proposition 27.** *Let $\mathcal{P}$ be the hyperplane of all $\mathbf{x} \in \mathbb{R}^k$ with $(\mathbf{x}-\mathbf{p})\cdot\mathbf{n} = 0$ such that $\mathbf{n} \neq \mathbf{0}$. Then the distance of any point $\mathbf{a}_x \in \mathbb{R}^k$ from the hyperplane $\mathcal{P}$ is given by*

$$\mathrm{dist}\left(\mathbf{a}_x,\mathcal{P}\right) = \frac{(\mathbf{a}_x - \mathbf{p}_i)\cdot\mathbf{n}}{\|\mathbf{n}\|} \tag{7.14}$$

with $\|\mathbf{n}\|$ the norm of the normal $\mathbf{n}$ and $\mathbf{p}_i \in \mathcal{P}$ is one of the Pareto points.

The final algorithm is illustrated geometrically in Figure 24. Even if we use the iSAX index, the implementation presented here can be used with any representation, distance and indexing techniques available (cf. Chapter 5). We simply assume that a time series index is constructed for each objective in order to perform efficient 1-NN queries and, therefore, avoid linear scan. We also consider that the index provides a lower bounding distance measure on indexing nodes (as explained in the previous section).

This implementation is presented in Algorithm 7.3. Given a query Q, a database `db` and a set of index `TS-Indexes` for each objective (constructed prior to the search), we start by transforming the query and computing the first-level distances as previously. This set $a\mathrm{Dist}$ is then used to perform the 1-NN exact queries on each objective. These queries give us the initial Pareto front $\mathcal{P}$ that form the approximate Pareto hyperplane. The 1-NN queries also compute a small portion of exact distances for each objective that we recover in list `aDist`. That way, after 1-NN queries we already have an approximate lower bounding position for each element. We then obtain the normal of the hyperplane defined by the list of Pareto points. Then, we evaluate each element of the database and stop distance computation as soon as they are dominated by the hyperplane. If we compute the complete distances in every objective for an element, we add it to the list of potential Pareto points. Finally, we filter this list by extracting the final Pareto front $\mathcal{P}$ at the end of the algorithm.

**Algorithm 7.3** MOTS matching algorithm by approximate hyperplane search.

```
multiobjectiveHyperplaneSearch(Q, db)
  // Quantify the query
  Q̄^{k∈[1...N_obj]} = { (w/n) Σ_{j=(n/w)(i−1)+1}^{(n/w)i} Q_j^k, i ∈ [1...w] }
  // Compute query-to-bin distances
  for k ∈ [1...N_obj]
    for b ∈ [1...N_bins^k]
      aDist_{i∈B_b^k}^k = D_approx(Q̄^k, B̄_b^k)
    end
  end
  // Perform efficient 1-NN queries
  [P aDist] = 1NN-Queries(Q, aDist, TS-Indexes)
  // Reference direction vector
  d = max_j(p_i^j), p_i ∈ P
  // Compute hyperplane normal
  n_p = ([P, d] [P, d]^T)^{-1} d
  // Transform into unit-norm vector
  n_p = n_p / sqrt(n_p^T n_p)
  for i ∈ [1...size(db)]
    for k ∈ [1...N_obj]
      if (aDist_i − p_1) · n_p < 0
        abandon
      else
         aDist_i^k = D_Q^k(S_i)
    end
    add(S_i, P)
  end
  checkParetoFront(pPoints)
```

## 7.4 EFFICIENCY ON MASSIVE DATABASES

We present the results of our algorithms regarding computation efficiency. Unfortunately, because of the novelty of this problem, there exist no competing method to compare. Hence, we are evaluating our methods against the *brute force multiobjective* algorithm on synthetic and real datasets. The artificial dataset is composed of random walk time series generated with a constant size of 512 time points. An independent set is synthesized for each hypothetical objective. The second (real) dataset is a combination of *Studio On Line* [23], *Real World Computing* [146] and *Vienna Symphonic Library* instrumental databases. These datasets include single notes of different playing modes from 23 orchestral instruments, which amounts to a total of 213.814 sound files. These files are WAVE and AIFF format, quantified to 16-bit at a sampling rate of 44.1 kHz. Subsets of the collections are randomly selected for increasing database sizes. Objectives are also randomly selected from a set of audio descriptors (cf. Table 2). This selection procedure is ten-folded. For each set of parameters (database size and set of objectives), one hundred queries are processed in order to avoid statistical anomalies. Queries are random walk time series with a constant size of 512 points. Computations were performed on a Macbook 2.4 GHz Dual Core running under Mac OS X 10.6.6 with 2 GO of DDR3 RAM.

### 7.4.1 *Comparing algorithms*

We present the results of different algorithms in terms of *querying wall time* for synthetic datasets in Figure 25. The left figure shows the *median* (dotted line), *average* and *variance* (solid line) in querying time for increasing database sizes. As we can see, the early abandon algorithm can already provide up to two times of speedup over the brute force approach, with a very low variance in the querying time. However, this factor of speedup appears to be linear to the cardinality of the dataset. The *hyperplane* algorithm is strongly superior, as it provides up to ten times of speedup over the brute force approach, and its median is usually even faster. The differences between the early abandon and hyperplane approaches can be explained by the higher number of Pareto front evaluations in the first one. However, the variance in querying times of the hyperplane search also increase with the cardinality, which imply that the resulting time might vary more importantly than early abandon depending on the distribution of the dataset. The most enthralling finding concerns the efficiency of our algorithm with increasing number of objectives, presented in Figure 25 (right). As we can see, the early abandon also provide a linear factor of speedup over the brute force approach. However, the hyperplane algorithm exhibits a sub-linear behavior when the number of objectives grows, once again with a significantly lower median. This sub-linear behavior could be explained by the higher probability that a large portion of the search space is ruled out by the approximate hyperplane with a greater number of dimensions.

To analyze this hypothesis, we compare the pruning power induced by each algorithm. The *space pruning ratio* is computed by comparing the proportion of points that are not entirely evaluated (ie. their complete distances are not calculated) to the quantity of points in the dataset. One of the main advantage of this measure is that it is hardware and dataset independendent, furthermore it is also independent of the complexity of the distance measure used in the final computation. Therefore, the gain provided by the different techniques can be compared objectively. Figure 26 (left) exhibits the space pruning ratio provided by the early abandon and the hyperplane al-

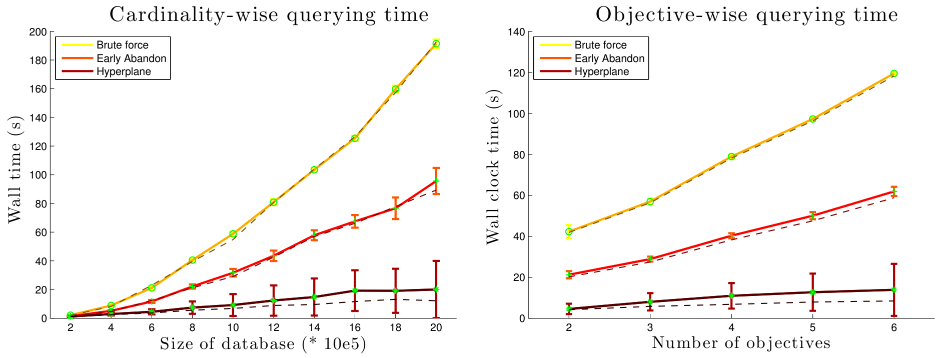



Figure 25: Query wall time (in seconds) for increasing database size (left) and increasing number of objectives (right) on synthetic datasets.

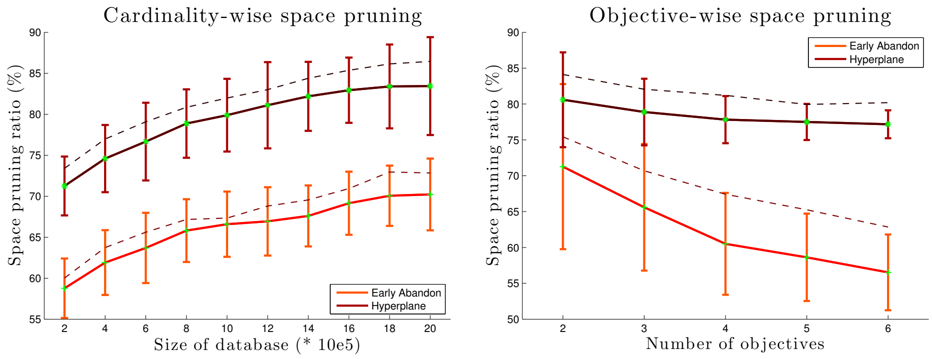



Figure 26: Space pruning ratio for increasing database size (left) and increasing number of objectives on synthetic datasets.

gorithms (the brute force is omitted as its pruning ratio is obviously null) for a growing amount of time series in the database. As we can see in this figure, the *hyperplane* limit provides a strongly superior pruning ratio as compared to the *early abandon* technique. The variances seem to remain constant (with small variations) for both algorithms as the number of objective grows, with once again a higher variance for the hyperplane algorithm. However, it must be understood that this measure is, in fact, a ratio of the number of elements evaluated. Therefore, an equivalent variance for higher cardinality will imply a higher variance in the number of elements pruned. In both case, the techniques seem to indicate an upper bound in pruning power as the number of time series increase. Figure 26 (right) exhibits the space pruning ratio provided for a growing number of objectives. As we can see, the hyperplane algorithm quickly converge to a constant pruning ratio around 80% (which can explain its sub-linear time complexity), whereas the early abandon algorithm seems to exhibit a continuous drop in pruning power.

### 7.4.2 *Comparing datasets*

We now compare the performances of different algorithms depending on the nature of underlying data. Even if the comparison on artificial datasets shows the strong superiority of the proposed hyperplane algorithm, a proper evaluation should rely

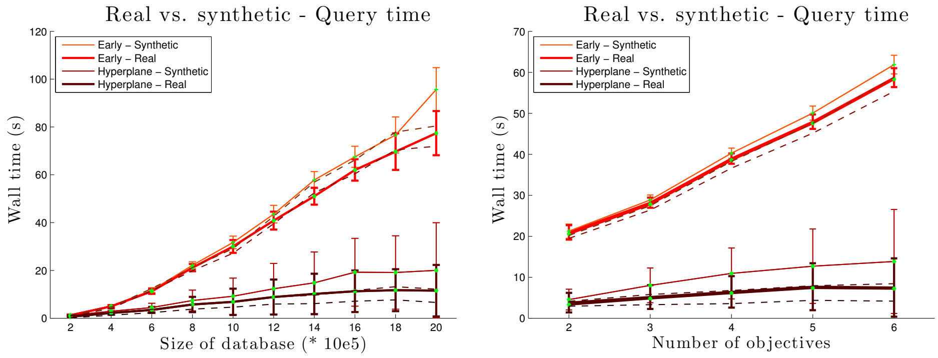



Figure 27: Query wall time (in seconds) for increasing database size (left) and increasing number of objectives (right) compared between synthetic and real datasets.

preferably on real datasets. Therefore, we analyze the efficiency of different algorithms on the audio collection presented previously and compare it to the results obtained on synthetic datasets. In order to achieve a meaningful comparison, the time series in the real dataset have all been resampled to a length of 512 time points. We present the results of algorithm *wall time speed* in Figure 27 for a growing number of series and objectives. We omit the results of brute force for the sake of clarity. Analysis of results reveals that both algorithms performs even better on real sound collections. This could be explained by the distribution of time series in real datasets, which is unlikely to be uniform as it is for random walk datasets. However, it seems that the enhancement is more pronounced for the hyperplane algorithm in both cardinality and objective-wise results. This seems to follow the intuition that the use of an approximate hyperplane benefits from the uneven distribution of data. Therefore, it enhances the overall efficiency of the algorithm.

We offer the same comparison for the space pruning ratio in Figure 28. As we can see, the higher performance for both algorithms is clearly seen in their respective pruning power for increasing databases sizes (left). The somehow more chaotic distribution of space pruning ratios can be explained by the nonuniform distributions of data, which cause a wider disparity in each evaluation. For an increasing number of objectives (right), the improvement seems to be a lot less noticeable, with the same overall evolution of pruning power with the number of objectives. In the real dataset, even if the pruning power of the hyperplane method is at first higher, its loss is slightly more substantial with a higher number of objectives.

## 7.5 INNOVATIVE AUDIO QUERYING

Now equipped with a flexible matching framework based on observation of the auditory perception, it seems logical to apply it to audio settings. We begin by showing the potential interest for our framework in content-based audio retrieval, by outlining the limitations in state-of-art audio matching methods (Section 7.5.1). We will describe the improvement gained by using the MOTS framework on the traditionnal problematic of *Query By Example* (QBE) (Section 7.6.1). Then, we will show how to go beyond this conventional approach and propose two innovative audio querying methods. These new paradigms can provide easier and more intuitive control over query specifications and at the same time bypass the need for a well-formed example.

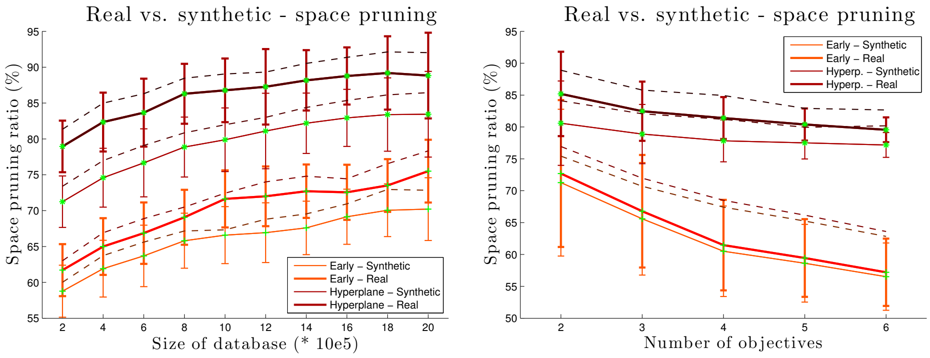



Figure 28: Space pruning ratio for increasing database size (left) and increasing number of objectives (right) on synthetic and real datasets.

### 7.5.1 *Content-based audio retrieval*

The past decade has witnessed a growing interest in *content-based retrieval* for multimedia databases [398]. Large amount of work has been devoted to performing intuitive queries over musical *songs* databases [77], such as the *Query By Humming* (QBH) approach [410], which is now a popular content-based music retrieval method. This paradigm allows finding a song in a large collection without knowing its name or artist, just by humming its melody. Tracing back to the seminal work of Ghias et al. [137], QBH systems typically rely on *symbolic* representations of melodies, rather than generic audio databases. Sound sample databases induce a greater challenge, as they are more massive and grow faster than musical databases. Furthermore, sound samples do not benefit from the same high-level symbolic information that can be extracted from melodies. Therefore, such sets may require an overwhelming amount of time to find a specific sample. The *Query By Example* (QBE) paradigm tries to tackle this problem by finding audio clips similar to a given sound example based on their spectral properties. The first QBE system was proposed by Wold et. al [380] where sounds were represented by a vector of spectral features, which were then compared with the Euclidean distance. This approach has subsequently been extended using larger sets of features [372] or other spectral transforms like the Discrete Cosine Transform (DCT) [342] and wavelet transform [222]. Several indexing and learning schemes have also been investigated like Nearest Feature Line (NFL) [223], Support Vector Machine (SVM) [153] or Gaussian Mixture Model (GMM) [165]. Other studies have focused on the temporal modeling of sounds, either by using templates of temporal energy [64] or Hidden Markov Model (HMM) [403] where comparison of HMM likelihoods with the query allows to obtain a ranked list of results. Finally, a different stream of generic audio querying is *Semantic Audio Retrieval* [332] which tries to discover the relationships between semantic and acoustic spaces. This provides queries on semantic concepts rather than acoustic features. This approach was implemented with a mixture of probability experts in [331] and extended with polysemy handling [68] and semantic weighting [27].

Generic audio retrieval is facing several problems that can be outlined from previous works in this field. First of all, metadata information is clearly inadequate to enable complex and intuitive interactions. It is almost impossible to maintain consistent and expressive metadata on large datasets. Semantic retrieval tries to provide a turnaround to manual annotation but still requires an extensively annotated starting set. Furthermore, it is limited to descriptive facts and sounds clearly related to a production source. However, most of the timbre 'qualities' cannot be captured using semantic concepts without subjective interpretation of data. Several authors pointed out the impossibility of sharing a common language for audio property description [110, 275]. This imposes severe limitations on the scope of possible queries, restricted to a predetermined set of semantic classes. Some QBE systems use clustering before retrieval, based on the idea that search time could be reduced by comparing the query only to a relevant cluster [164, 344, 404]. However, building hierarchical classes implies that the database is created according to a particular dataset. Therefore, once the database is built, it loses flexibility and users have to adapt to this original hierarchy. Finally, as we discussed earlier, authors have pointed out the multifaceted perception [110] and the unlikeliness for a single measure of perceptual similarity of audio signals [363]. Therefore, sound retrieval systems should be flexible enough so that variable influence could be put on different sound properties during perceptual similarity evaluations [245], but yet no current audio-retrieval system seems to address these limitations.

### 7.5.2 *Going beyond traditional query paradigms*

We now show how the application of the MOTS paradigm allows us to handle previously listed problems. Our approach relates to [275] where sounds with or *without a known cause* are described by looking specifically at the temporal evolution of their acoustical properties. Sound clips are considered as short-duration *units of musical creativity* [76]. In order to maintain the flexibility of the database, we avoid the clustering paradigm by deliberately not interpreting data. Therefore, no assumptions are made on spectrum types and sounds can be of any nature. In order to provide more comprehensive query conditions, we do not use semantic annotation and focus on the temporal evolutions of timbre properties which provide objective comparisons.

First, our system is based on time series data mining method (cf. Chapter 5), relying on a pre-constructed database structure (that we detail in Section 7.6). This system alone could already provide a possibility of high-level sound querying by directly matching the temporal evolution of each spectral descriptor. This would allow to find sounds by simply drawing the desired evolution of a timbre characteristic. However, as we discussed earlier, we want to confront explicitly the multidimensionality of timbre perception (cf. Section 3.3) for multiple audio features input. This paradigm shift is presented in Figure 29. In the QBE approach (left), a soundfile is fed to the system for which similar sounds have to be found in a database. The system answers with an ordered list of soundfile results. By using the time series techniques (center), we can construct a system which match the temporal evolution of any audio feature. However, the combination of several audio features input (right) requires a more flexible matching process, hence exhibiting the relevance of the MOTS framework.

Hence, we further consider a core problem of audio retrieval that lies in the query specification itself. As put forward by Donwie [110], audio queries are themselves a form of musical information, and are, therefore, complex and multifaceted. Several authors pointed out that most users of content-based retrieval systems have only a vague idea of what they seek at the onset [220, 377, 387]. Hence, they might also search for aspects of an audio query but not exactly the same content. We will show how the MOTS results handle this aspect by being presented in an informative way to users. Finally, when an example is unknown or difficult to generate, the query should help the user determine what he is seeking by being specified in a manner

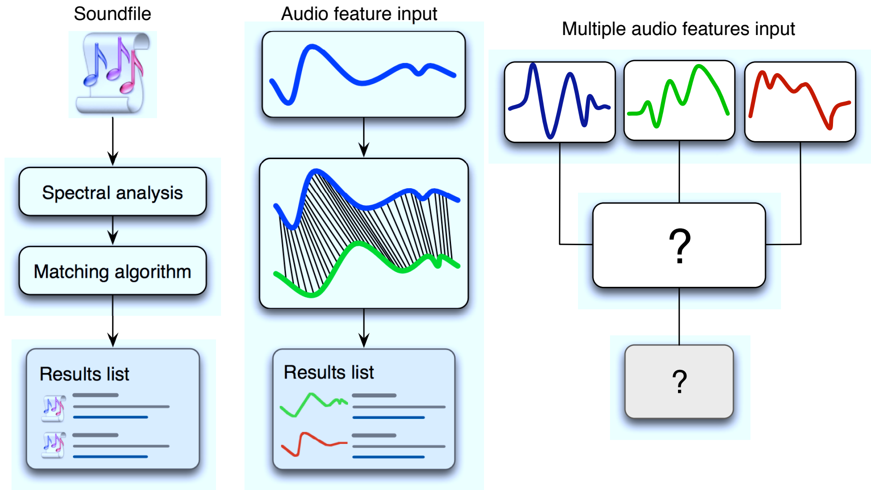



Figure 29: Shifting from the QBE paradigm (left) to the MOTS framework (right). In the QBE approach (left), a soundfile is fed to the system for which similar sounds have to be found in a database. The system answers with an ordered list of soundfile results. By using time series techniques (center), we can construct a system which match the temporal evolution of any audio feature. However, the combination of multiple audio features input (right) requires a more flexible matching process, hence exhibiting the relevance of the MOTS framework.

as close as possible to the underlying nature of audio properties [291]. To address all these shortcomings, we present two novel paradigms for audio querying based on the MOTS approach. First, the *MultiObjective Spectral Evolution Query (MOSEQ)* provides a flexible query specification by allowing users to draw directly schematic temporal shapes required for spectral features. Therefore, it bypass the need for a well-formed example. Based on this paradigm, we introduce the *Query by Vocal Imitation (QVI)*, which allows users to perform vocal imitations of desired properties. In both cases, the system returns the samples shown on a multidimensional front depending on how well they match the different time-evolving timbre dimensions, thus providing flexibility in results representation. Furthermore, equipped with the MOTS framework and adequate similarity measures along each perceptual dimension, we are able to predict various degrees of similarity between elements. Figure 30 summarizes the algorithmic framework for both applications.

## 7.6 DATABASE STRUCTURE

As we perform queries over large collections of sound samples, we have to maintain a structured database. Figure 31 depicts how sounds are analyzed and managed. We process sound samples with IRCAMDescriptor [274] in order to extract all perceptually relevant information from low-level signal data. The list of descriptors is provided in Table 2.

The mean and standard deviation of each descriptor are extracted and stored in the database. We then normalize the temporal shapes in order to obtain *zero-mean* and *unit-variance* time series. We then store the entire time series, using the SAX representation [231]. Therefore, each element in the database contains several time series which represent different characteristics of a sound. The temporal shapes are

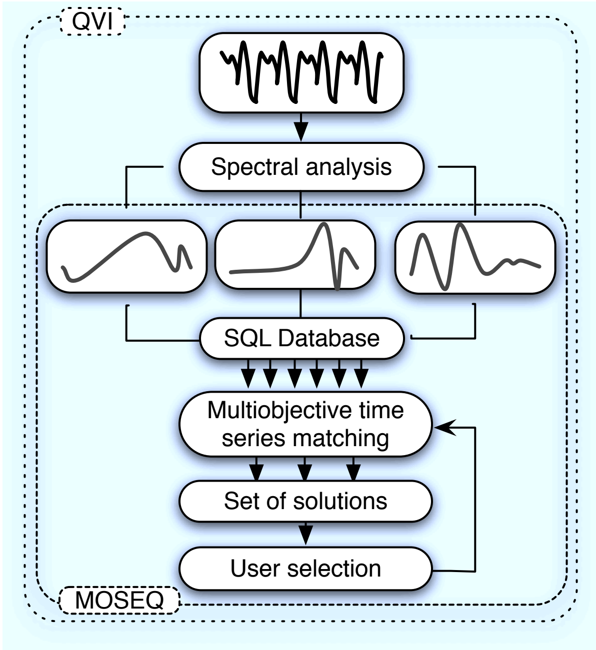



Figure 30: Algorithmic framework for two types of interaction. In *MultiObjective Spectral Evolution Query* (*MOSEQ*), a set of time-evolving properties is drawn. The *MOTS* algorithm allows to find the set of efficient solutions. In *Query by Vocal Imitation* (*QVI*), the user can directly use his voice to perform an imitation of the desired properties. A spectral analysis leads to the set of properties.

| **Category** | **Features** |
|---|---|
| Energy | *EnergyEnvelope, HarmonicEnergy, Loudness, NoiseEnergy, TotalEnergy* |
| Spectral | *FundamentalFrequency, Inharmonicity, Noisiness, Sharpness, Spread, Flatness, Crest, Centroid, Skewness, Kurtosis, Slope, Decrease, RollOff, Variation* |
| Harmonic | *Deviation, OddToEvenRatio, Tristimulus, Centroid, Spread, Skewness, Kurtosis, Slope, Decrease, RollOff, Variation* |
| Perceptual | *Roughness, Deviation, OddToEvenRatio, Tristimulus, Centroid, Spread, Skewness, Kurtosis, Slope, Decrease, RollOff, Variation* |
| Sub-bands | *MFCC, RelativeSpecificLoudness, AutoCorrelation, Chroma, ZeroCrossingRate* |

Table 2: List of available descriptors whose mean, deviation, temporal shape and first and second derivatives are stored separately. More detailed information can be found in [274]

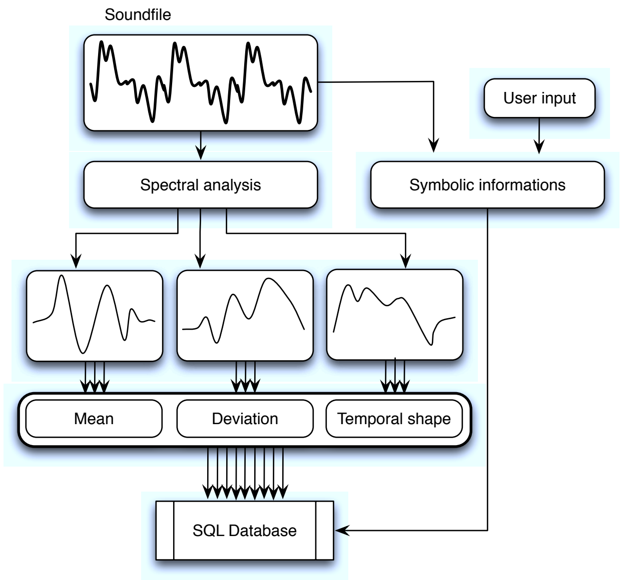
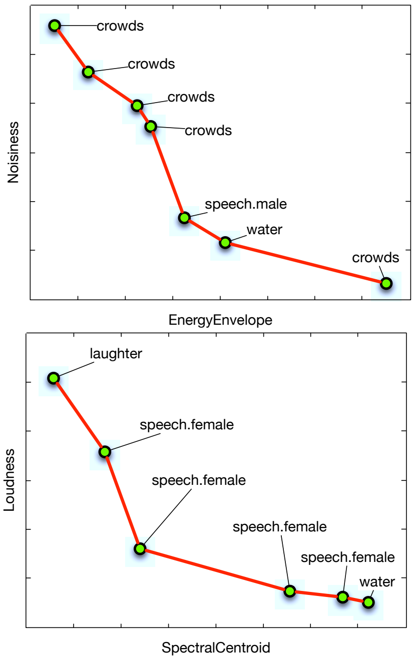



Figure 31: When a soundfile is input to the system, the analysis module computes a set of descriptors whose mean, deviation and temporal shape are stored separately inside an SQL database. Symbolic information can also be stored in the database, either by automatic extraction or direct user input.

resampled to a uniform length. This could be considered as a concern for audio querying as long sounds can be compared to extremely short sounds. However, this approach shows the benefit to focus solely on the temporal shape. Furthermore, the system allows using duration in conjunction with other objectives to be optimized. The length can alternately be defined as a filtering constraint which will reduce the search space to sounds of matching length. Other symbolic information can be added to the database, either by automatic extraction from filenames or direct user input. However, we consider in the final search problem that no metadata is available whatsoever.

### 7.6.1 *QBE results and representation*

As our approach is multiobjective, query results are presented as a Pareto front in a multidimensional space. Figure 32 present the results of two queries with the MuscleFish dataset (we detail this dataset in Section 11.1.2, as it will be used to validate the MOTS framework on classification tasks). The first query (left) is performed using a restaurant scene belonging to the *crowds* class. The second query (right) is performed using a sample of *female speech*. For each query, we compare the results of mono-objective and multiobjective methods given the same set of features. Mono-objective selection provides an ordered list of results. However, there is no informed knowledge about how these choices were made whatsoever. Even with multiple dimensions involved, the results only offer an "*optimization line*" of fitness. Oppositely, the multiobjective framework allows to obtain the complete optimization space. This representation informs the user on how solutions optimize various objectives. It also allows users to explore this space by focusing more on one objective than the other. An obvious limitation of this system is on the number of dimensions that can be used for representation. However, we can easily represent three-dimensional cuts of any multidimensional space. If we




| | |
|---|---|
| 1 | speech.male |
| 2 | water |
| 3 | speech.male |
| 4 | bells |
| 5 | crowds |
| 6 | water |
| 7 | water |
| 8 | animals |
| 9 | water |
| 10 | speech.female |

| | |
|---|---|
| 1 | speech.female |
| 2 | oboe |
| 3 | percussion |
| 4 | cellobowed |
| 5 | percussion |
| 6 | percussion |
| 7 | percussion |
| 8 | speech.female |
| 9 | machines |
| 10 | violinbowed |

Figure 32: Comparison of different query results for multiobjective optimization and mono-objective selection in a QBE context. (Left) A sound taken from a restaurant scene and belonging to the *crowd* class. (Right) A clip taken from the *female speech* class.

look more closely at the results of these queries, we can see that the sets provided by the MOTS approach are more similar to the initial example query. In the first case, it appears that relevant results are spread over the criteria space. This distribution is revealed by the multiobjective matching. On the other hand, mono-objective selection seems to get stuck on solutions performing averagely in both objectives. Furthermore, by separating every optimization dimension, the MOTS representation already entails the case where users seeks parts of the query but not exactly the same content with multiple definitions of similarity. Even if this first evaluation remains an extremely narrow and empiric trial, we will discuss the validation of the MOTS approach for audio querying through extensive user studies in Section 7.7 and complete classification tasks in Section 11.1.

### 7.6.2 *MultiObjective Spectral Evolution Query (MOSEQ)*

We present our first application of the MOTS framework that provide a novel way to find sounds in massive databases. The MOSEQ paradigm allows users to draw directly the temporal evolution of several audio features to be found in the database. Therefore, it bypass the necessity of a well-formed example and also provide a flexible query specification. We start by introducing the definitions required for this application.

*Definitions*

**Definition 28.** *Sound attributes.* An attribute $\mathcal{A}$ is a symbolic value representing meta-informations about a sound in the database. For instance, if $\mathcal{A}_{dyn}$ is the dynamics attribute, then $\mathcal{A}_{dyn}(\mathcal{S}) \in \{pp, p, mf, f, ff\}$. Sound attributes are obtained by manually tagging the sample database.

These may be used to define symbolic constraints on the samples and thus restrain the size of the search space. However, we do not consider symbolic attributes in the final search problem.

**Definition 29.** *Sound features*. A feature $\mathcal{F}$ is a numerical (eventually multidimensional) value that describe perceptual aspects of samples. Sound features are extracted from the raw signal data (cf. Section 7.6) and can represent temporal evolutions or mean descriptors. In the latter case, they are used as attributes for reducing the search space.

We will thereafter consider the computer scientist view on timbre (cf. Section 2.3.3), which suggests that any timbre may be fully characterized by a set of sound features $\{\mathcal{F}^1, ..., \mathcal{F}^K\}$.

**Definition 30.** *Target*. The target query $\mathcal{T}$ is the timbre that is input to the system and for which similar instances are to be found in the database.

For the MOSEQ system, the target is, therefore, represented by a set of time series features $\{\mathcal{F}^1(\mathcal{T}), ..., \mathcal{F}^K(\mathcal{T})\}$ which can be of variable cardinality $K \in \{1, \ldots, \mathcal{N}_{features}\}$

**Definition 31.** *Features similarity functions*. Given a sound target $\mathcal{T}$ represented by its feature set $\{\mathcal{F}^1(\mathcal{T}), ..., \mathcal{F}^K(\mathcal{T})\}$ and a sound sample $\mathcal{S}$, the $k^{th}$ similarity function is the real-valued function $\mathcal{D}^k_{\mathcal{T}}(\mathcal{S})$ that returns the distance between a sound $\mathcal{S}$ and the target $\mathcal{T}$ along the $k^{th}$ feature. In other words, $\mathcal{D}^k_{\mathcal{T}}(\mathcal{S})$ is a similarity measure between time series features $\mathcal{F}^k(\mathcal{S})$ and $\mathcal{F}^k(\mathcal{T})$.

*Paradigm*

The idea behind this interaction paradigm is that when a user seeks a sound sample, he will create a mental representation of the corresponding sound based on the temporal evolution of several spectral properties. Therefore, the MOSEQ system allows to draw these shapes in order to project a mental representation into an efficient query. The user selects a set of features that are relevant to his query. For each, he can draw the desired time series. This set acts as the target for the system. Therefore, it bypass the need for a well-formed example. We consider that the database follows the structure described in Section 7.6 and contains several sound features $\mathcal{F}^i$ for every sample. In a QBE context, the target $\mathcal{T}$ is the sound example for which similar instances have to be found. For the MOSEQ system, the target is represented by a set of time series features $\{\mathcal{F}^1(\mathcal{T}), ..., \mathcal{F}^K(\mathcal{T})\}$. As noted earlier, given this target and a sound sample $\mathcal{S}$, each of the similarity functions $\mathcal{D}^k_{\mathcal{T}}(\mathcal{S})$ can be defined using a different function for each objective. Therefore, the goal of the MOSEQ paradigm is to optimize simultaneously the entire set of time series features sought by the user. By using the MOTS approach, the system display the multidimensional space containing audio clips that jointly optimize the sound features. We will discuss the validation of this paradigm in Section 7.7 through extensive user studies.

### 7.6.3 *Query by Vocal Imitation (QVI)*

Now equipped with the MOSEQ paradigm, we can go even further in terms of ease of interaction. In several context, the most straightforward way to communicate a musical idea is to use one's voice. Therefore, we believe that a natural way of querying sound samples would be to directly input a vocal imitation as a query. Indeed, the vocal system can produce a wide variety of sounds. Most people have in some occasions imitated everyday sounds by using their voice and presumably tried to match the temporal evolution of acoustic properties. For musicians, the use of nonsense text singing, called *syllabling* [345], is an effective communication language for pedagogical purposes. Even with the inherent limitations of human voices, such as our frequency range *(tessitura),* we can control several vocal disorders. The *growl* effect increases expression by producing a rough sound. The *breathy* effect allows to generate noisier sounds. We can even learn to control extremely specific sound qualities like the position of the formant frequencies, the type of phonation or the singer's formant [346]. We thus benefit from the high degree of expression of the singing voice, principally described by loudness, fundamental frequency and spectral envelope, which all vary dynamically with time. However, it is obvious that the capabilities of a human voice are inherently limited in terms of spectral features. Nevertheless, sung imitations may convey valuable information as Pressing [287] indicates: "*One important resource in designing such expressivity is to use one's own voice to sing the expression in the part. Even if the sound quality is beyond the powers of your (or perhaps anyone's) voice, its time shaping may be imitable*". Therefore, despite the voice is limited in the range of timbres it can produce, much of vocal expression can be captured, not in the absolute timbre but in the relative temporal variation of timbre. This exhibit that our intuitions on temporal evolution can be appropriate in this setting, and we will discuss the evaluation of the efficiency and usability of the QVI framework through extensive user studies in Section 7.7.

Therefore, the QVI problem is defined as processing a vocal imitation provided by the user. This imitation is input to the same analysis module used for filling the database. Therefore, it provides the spectral shapes within which the user can select its desired criteria. Hence, the QVI problem can be reduced to a MOSEQ problem but still provides an equivalent of QBH for sound samples. The MOTS framework is then used to optimize the various features selected by the user jointly. We now show how to combine both paradigms and provide turnarounds to support the lack of vocal control capabilities over spectral properties.

*Combining paradigms*

Despite the unavoidable limitations of one's voice, it is still possible to go beyond these with two possible turnarounds. First, missing dimensions can be specified by manually drawing the desired curves of evolution. The problem then turns to be a MOSEQ which is partly defined manually. Second, a standard mechanism used in the interaction field is to convert a control signal into another parameter (a procedure known as *mapping*). Vocal range can easily be mapped to any target range by transposing the incoming voice pitch. Hence, it is also possible to establish a mapping between any useful vocal descriptor and unrelated spectral dimensions.

## 7.7 EXPERIMENTAL VALIDATION

The paradigms proposed in the previous section strongly rely on the premise that we are able to perceive and create mental representations of high-level audio features. The last decade of research have established a wealth of knowledge on the main properties of audio signals (namely *loudness* and *fundamental frequency*). However, the perception of more complex audio features and their temporal evolution has yet to be investigated. Furthermore, we hypothesized on the intuitivity and ease-of-use of the querying paradigms. Hence, it appears mandatory to validate the MOTS framework through perceptual studies and the MOSEQ and QVI paradigms through usability evaluations. A first step towards the evaluation protocols have been formalized during a research exchange project with the CIRMMT laboratory of McGill university in Montreal, under the supervision of Stephen McAdams.

In this experimental protocol, we focus on studying the multidimensional perception of the temporal evolution of higher-level sound features. We start by trying to see how well temporal variations of these features can be perceived. We then study the relative perceptual importance of the features when compared to each other in a setting where they are maximally decorrelated. By relying on the concept of *directions of listening*, we try to access to the structure of our multi-dimensional space of perception. We study the consistency of these directions of listening through several tasks of direct similarity, generic similarity and shape drawing. Finally, we take advantage of this multidimensional framework to perform an extensive usability evaluation of the two innovative audio querying. In this context, we must face experimental constraints coming from both fields of user interface (UI) and audio-related evaluations which induce large levels of variability. Therefore, we decided to conduct this study by using the *Usability Evaluation* (UE) Ellis and Dix [118] framework, which allow us to draw knowledge from the researches conducted in UE over the past years. *Usability* is the extent to which a system enables users, to achieve specified goals *effectively* and *efficiently* while promoting feelings of *satisfaction* (ISO 9421-11 [1]). We can see that this definition applies especially well in the context of quering systems.

This experimental evaluation is currently being simultaneously assessed at the CIRMMT laboratory in Montreal and the IRCAM laboratory in Paris. The complete protocol and experimental results are presented in the Annex **??** of this document.

# Part IV

# HYPERVOLUME CLASSIFICATION (HV-MOTS)

# 8 HV-MOTS CLASSIFICATION

As we have just seen in the previous chapter, the MOTS approach allows to find the set of efficient solutions in a database given multiple time series features. We showed that this approach gives access to more flexible sets of solutions in the context of retrieval and querying. However, we could wonder if this flexibility can also benefit to other field of studies. Notably, as the MOTS framework is intended to find efficient sets of similarities in a database, it fits the basic requirements of classification paradigms. However, given the definition of Pareto optimality, there is normally no way of ranking the different elements between each other. Therefore, there is no straightforward criterion to make a final classification decision. We introduce the notion of *hypervolume dominated* by a class and show how to use it as a classification criterion. Our main idea is that by not merging every dimensions into a single distance measure, we can benefit from a more accurate and flexible view on the properties of various classes. We will show that using this multi-objective flexibility with the adequate hypervolume criterion allows to construct a classifier that significantly outperforms state-of-art methods. Specifically, our approach exhibits a statistical superiority over the 1NN-DTW classifier which has been consistently shown to still be the best performing classification scheme for time series [149, 150, 177, 293, 292].

## 8.1 MULTI-OBJECTIVE CLASSIFICATION

Based on the MOTS selection, we still need a criterion to make the final classification decision, ie. to select which class is the best match to a given input. We introduce in this section three class selection criteria that allow to use the Pareto optimality and apply it to any classification problem. We consider that elements to be classified are compared to known items thanks to different distance measures on several dimensions. We also consider that these distances are not merged and give access to a complete multidimensional distance matrix.

### *Pareto cardinality*

Given the Pareto set, we can first simply look at its cardinality. Therefore, our first selection criterion can be obtained by counting the number of occurrences of each class $c$ in the Pareto front $\mathcal{P}$. The selected class $\mathcal{C}_s$ is the most represented in the front.

$$\mathcal{C}_s = \underset{c}{\mathrm{argmax}} \left( \left| p_i \in \mathcal{P}, \mathrm{class}\left(p_i\right) = c \right| \right) \tag{8.1}$$

This is obviously a simple criterion and we can expect it to be less efficient in higher dimensions. We term this method *MOTS* in the following.

### *Nearest Pareto*

Given an input, each class $c$ can provide a different Pareto front $\mathcal{P}_c$ depending on its distances sub-matrix. Therefore, our second selection criterion can obtained by computing the Pareto front of each class and then computing the mean distance between the input and all elements in the front. The selected class is therefore the one which implies the minimal distance.

$$\mathcal{C}_s = \underset{c}{\operatorname{argmin}} \left( \frac{1}{n} \sqrt{\sum_{i=1}^{n} \|p_i\|^2}, p_i \in \mathcal{P}_c \right) \tag{8.2}$$

We term this method *Nearest Pareto MOTS (NP-MOTS)* in the following.

*Hypervolume domination*

We now introduce a novel criterion based on *hypervolume* domination. This measure has been used in multi-objective optimization with Genetic Algorithms (GA) [411] as a performance indicator, i.e. only to differentiate the quality of different algorithms. However, to our best knowledge, it has never been used as a classification criterion. The idea behind this measure is that every point in a multi-dimensional space, defines a hypervolume which indicates the portion of space Pareto dominated by this point. For a $n$-dimensional space, $n \in \mathbb{N}$, the hypervolume of a box in $\mathbb{R}^n$ generated by two points $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ is defined as

$$\mathcal{H}(B) = \prod_{i=1}^{n} (b_i - a_i) \tag{8.3}$$

The *hypervolume dominated* by a Pareto front $\mathcal{P}$ given a reference point $r_p = \left(r_p^1, \ldots, r_p^n\right)$ is given by the union of hypervolumes dominated by each point in the front

$$\mathcal{H}(\mathcal{P}) = \mathcal{H}\left(\bigcup_i B_i\right) = \mathcal{H}\left(\bigcup_{(p_1,\ldots,p_k)\in\mathcal{P}} [p_1, r_p] \times \ldots \times [p_k, r_p]\right) \tag{8.4}$$

These notions are shown in Figure 33 (left). Point $p_1$ defines a box $\mathcal{B}_1$ (darker gray) with the reference point $r_p$. Each point of this set also implies a corresponding domination box. The hypervolume dominated by the Pareto front is therefore the union of hypervolumes dominated by each point in the front. Figure 33 (right) shows the benefits of this measure when comparing two distributions. Even though the first class have more elements belong to the final Pareto front, its dominated hypervolume $\mathcal{H}_1$ is smaller than the hypervolume $\mathcal{H}_2$ of the second class. Therefore, the hypervolume indicates both the fitness of a distribution and its spread over the optimization space. Furthermore, compared to a NN or NC rule, it summarizes the behavior of the whole class with respect to the input rather than the position of the input relative to a single element of the classes. In our implementation, we use the hypervolume computation algorithm proposed by [125]. We compute the hypervolume dominated by the Pareto front of each class. The selected class is therefore the one which induces the largest dominated hypervolume.

$$\mathcal{C}_s = \underset{c}{\operatorname{argmax}} \left(\mathcal{H}\left(\mathcal{P}_c\right)\right) \tag{8.5}$$

We term this approach *HyperVolume-MOTS* (HV-MOTS) in the following.

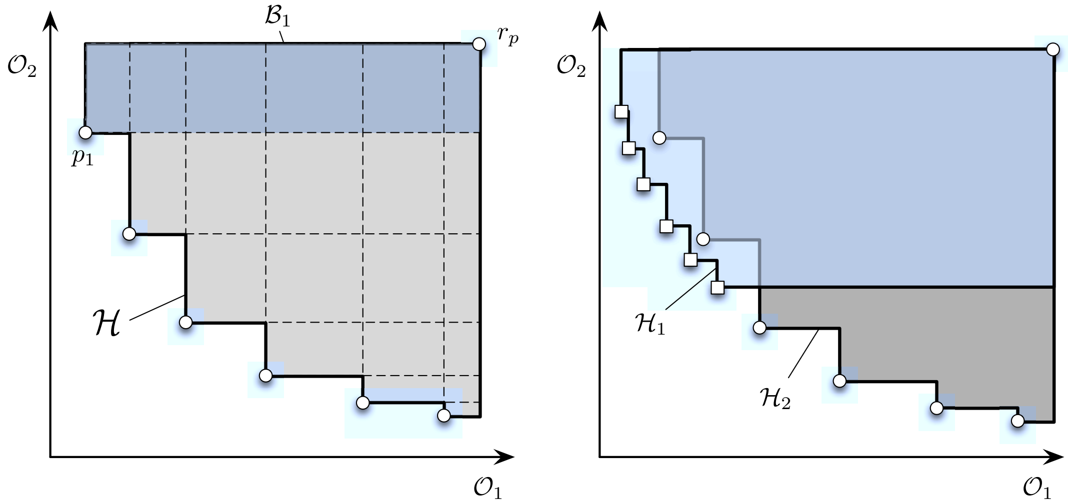



Figure 33: (Left) Hypervolume dominated by a Pareto front given the reference point $r_p$ in a 2-dimensional space. The darker gray subpart defines the box $\mathcal{B}_1$ which is dominated by point $p_1$. The hypervolume $\mathcal{H}$ dominated by the Pareto front is defined as the union of all boxes dominated by each point of the front. (Right) Comparison of two dominated hypervolumes $\mathcal{H}_1$ and $\mathcal{H}_2$. Even though the first class have more elements belong to the final Pareto set, its hypervolume $\mathcal{H}_1$ is smaller than the hypervolume $\mathcal{H}_2$ of the second class.

## 8.2 DISTANCE-BASED CLASSIFIERS

The HV-MOTS classification framework falls in the category of distance-based classifiers. To see the novelty implied by our proposal, we start by comparing it to other distance-based classifiers. Figure 34 illustrates these concepts. The element to be classified is represented by the cross at the origin of the space. The *Nearest-Neighbors* techniques will try to find the nearest element(s) based on the norm of the distance vector, thus defining the selected class accordingly. The *Nearest Center* technique performs the same analysis but by first computing the centroid of each class and then selecting the nearest one. The *MOTS* paradigm computes the Pareto front and then selects the class that is the most represented inside the set. The *NP-MOTS* paradigm first computes the Pareto set of each class and then selects the nearest one based on their mean distances. Finally, the *HyperVolume-MOTS* technique computes the hypervolume dominated by each class and then selects the class with the largest hypervolume.

As we can see in this figure, unlike the other classification schemes, the HV-MOTS technique does not perform a linear merging of the distance measures in every objectives. Instead, the computation of the hypervolume allows to account for two aspects of the class distributions. First, the *proximity* (as for the other schemes) is implied, given that the hypervolume will grow if the elements of the front are closer to the input. However, this criterion also accounts for the *spread* of the classes over the distance space. This therefore allows for flexible matching over both dimensions separately. Hence, it analyzes if classes to perform well in *every* directions of optimization.

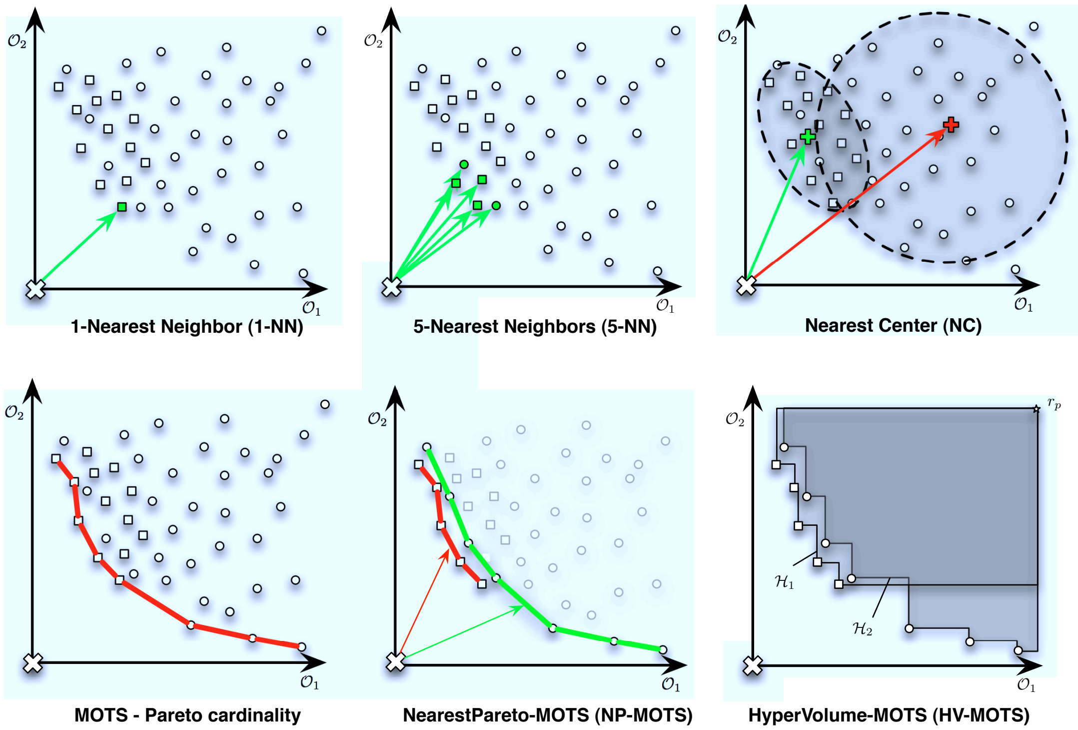



Figure 34: Comparison of distance-based classifiers. The element to be classified is represented by the cross at the origin of the space. The *Nearest-Neighbors* techniques select the class of nearest elements based on the norm of their distance vector. The *Nearest Center* technique first computes the centroid of each class and then selects the nearest one. The *MOTS* paradigm computes the Pareto front and then selects the most represented class. Finally the *HV-MOTS* technique computes the *hypervolume dominated* by each class and then select the class with the largest one.

## 8.3 COMPARISON TO OTHER CLASSIFIERS

We provide here a brief comparison on the class boundaries provided by the HV-MOTS classification framework, as opposed to other well-studied state-of-art classifiers. We separate this comparison between the methods which provide *linear* or *non-linear* class boundaries. We underline that this comparison is far from exhaustive but we selected a subset of classification methods that will be used in the subsequent large scale analysis of the HV-MOTS classifier. Figure 35 illustrates the comparison of several classifiers including our approach in terms of class boundaries in *feature* space.

NEAREST-CENTER As discussed previously, the *NC* classifier works by finding the center of class distributions and, therefore, the most likely positions of their elements. Then the input is compared only to these centroids. As we can see in Figure 35, the corresponding boundary is *linear* as the distance function delineates the boundaries between class pairs.

NEAREST-NEIGHBOR The *1-NN* approach tries to find the element that is closest to the input in terms of features distances, usually through their Euclidean norm. Hence, as it compares an input to every element of a class, the reference point changes

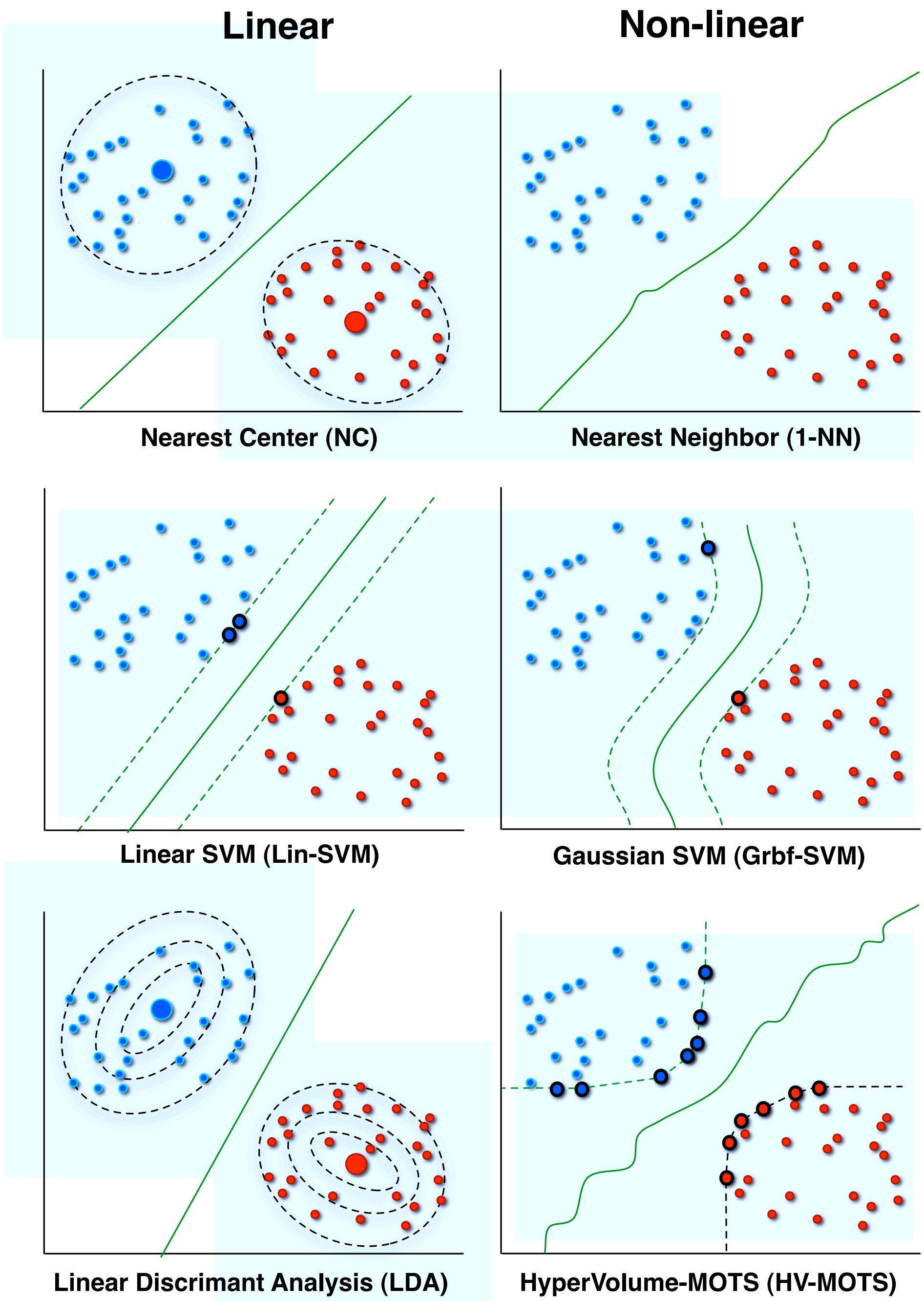



Figure 35: Comparison of several classification approaches based on the class boundaries that they define in *feature* space. The techniques are seperated between their definition of *linear* (left) or *non-linear* (right) class boundaries.

depending on the input. Therefore, as we can see in Figure 35, the resulting boundary is *non-linear*.

SUPPORT VECTOR MACHINES The *SVM* classification is based on two key ideas. First, it relies on the notion of *maximal margin* between the class boundary (solid line in Figure 35) and the closest elements to it. This gives a second boundary (dotted line in Figure 35 ), which defines the *margin* for decision. Hence, the elements selected for this margin are called *support vectors*. The second notion lies in the concept of *kernel*. This idea is introduced to solve the classification cases where the different classes are not linearly separable. Therefore, the kernel is applied to transform the data into a higher dimensionality where the separability is linear. We detail two possible kernels for SVM.

**Linear kernel** The simplest case of SVM is to use a linear kernel (ie. no dimensionality modifications are applied). This allows to position a decision hyperplane between the classes. The algorithm selects the hyperplane that provides the maximal margin (maximum separation) between the classes. As we can see in Figure 35, the Lin-SVM provides a *linear* separation between classes.

**Gaussian kernel** If the data requires a non-linear separation, a kernel function can define an alternate high-dimensional space. The main idea here is to use a set of functions called *Gaussian Radial Basis Functions* (GRBF), which are applied to the datas. The resulting space allows to use a linear separation based on the same hyperplane selection. Figure 35 exhibits the corresponding *non-linear* class boundary, maximal margin and support vectors.

LINEAR DISCRIMINANT ANALYSIS The *Linear Discriminant Analysis* (LDA) tries to find the dimension which maximize the inter-class variance and intra-class coherence. The ratio between these two values (sometimes called *inertia ratio*) is maximized in order to find the *linear discriminant* dimension. The data is then projected on this dimension in order to place a classification threshold between the classes means. This allows to obtain a hyperplane boundary between classes which is orthogonal to this dimension. As we can see in Figure 35, this leads to a *linear* class boundary.

HYPERVOLUME-MOTS We try to provide the main differences and ressemblances between our proposal and the presented classifiers. First, the HV-MOTS approach works in a fashion similar to NN and NC by trying to select the most similar class directly from the distance matrix. However, our approach is fundamentally different from these two methods, as it does not use a "direct distance" approach between elements. Indeed, the HV-MOTS scheme studies the overall distribution of each class separately. Therefore, it allows to analyze both the *distance* and *spread* of various classes. Moreover, by not using an Euclidean norm for elements-to-input distance and not merging dimensions into a single measure, each of the dimensions are treated separately. HV-MOTS also falls into the category of non-linear class boundaries such as NN and Grbf-SVM. Compared to the NN approach, however, we can see in Figure 35 that the delimited boundary between classes is somehow more complex. This comes from the fact that it uses information of the density and distance on dimensions separately. However, it is interesting to note here that for a single dimension, HV-MOTS is strictly equivalent to 1-NN classification. Indeed, as only one dimension is involved, the largest hypervolume is necessarily induced by the nearest element.

## 8.4 DISCUSSION

### 8.4.1 *Past results*

In the context of time series classification, it has been repeatedly proven that the 1-NN classifier is extremely hard to overpower [382]. Several published classification studies confirms that 1-NN classification is still the best performing classification scheme for time series retrieval [149, 150, 177, 293, 292]. Some authors even point out that "*while there have been attempts to classify time series with decision trees, neural networks, Bayesian networks, support vector machines etc., the best published results (by a large margin) come from simple nearest neighbor methods*" [106]. Even the SVM which is one of the most powerful classification scheme available appears to be *at most* statistically equivalent to 1-NN but usually performs worst [149]. This comes from the fact that time series are inherently high-dimensional data. Therefore, it is very difficult to define statistical methods that could avoid this *curse of dimensionality*. Therefore, the best performing methods are those based on distance comparisons.

As it has been repeatedly shown (in time series classification) that the 1-NN classifier outperforms other classification approaches, we will focus on comparing our approach to this framework. We will show in the large scale study that our novel approach statistically outperforms the 1-NN selection scheme.

### 8.4.2 *Advantages and drawbacks*

We try to discuss here the theoretical advantages and drawbacks of the HV-MOTS classification framework. As this method is a distance-based classifier, it provides similar disadvantages. First, as the NN classifiers, an obvious disadvantage is the time complexity of making the final decision. Indeed, the distance-based methods requires to compute the complete distance matrix. For the HV-MOTS selection, we can use the algorithms introduced in Section 7.3 that provides a significant speedup to find the Pareto front of each class. However, this imply to construct a different index for each class.

A major advantage of the HV-MOTS method is that *no assumptions* are made on the data or its distribution. Therefore, it can classify any type of data. Furthermore, as other distance-based methods, *no training* is required by this approach. Therefore, there is no need for extensive statistical models to be constructed beforehand. Another advantage of distance-based methods is that they are not bound by the *curse of dimensionality* involved in time series matching. Therefore, through the use of pairwise distances, the underlying data can be of any complexity without requiring the application of dimensionality reduction techniques. Furthermore, the HV-MOTS approach requires *no parameters* for classification, which makes it applicable to any dataset in a straightforward manner.

Finally, it has been shown that one of the main problem of $k$-NN techniques is that it tends to be very sensitive to irrelevant features as every dimensions contribute to the classification. A way of avoiding this problem is to weigh the different dimensions [95]. Therefore, compared to 1-NN selection, this is the main enhancement provided the HV-MOTS approach. Indeed, it specifically targets this problem and allows to alleviate its drawbacks, by considering *every possible weighting* in 1-NN selection. This theoretical improvement is straightforward from the fact that each element in the Pareto front is the best 1-NN selection given a particular set of weights.

### 8.4.3 *Comparison*

We try to provide here a deeper comparison of the HV-MOTS classification and 1-NN selection based on the class boundaries induced by both methods in feature space. Figure 36 illustrates the comparison of boundaries between 1-NN and HV-MOTS on two synthetic sets of data. The first problem (up) represents synthetic data where the classes are *almost* linearly separable. However, a slight intersection of class properties appears at the boundary. As we can see, the 1-NN defines harsh boundaries around these areas, by only taking into account the proximity of *nearest* elements. Oppositely, the HV-MOTS approach also accounts for the distribution of other points in the surroundings. Therefore, the overall boundary defined by HV-MOTS seems to be smoother than the 1-NN boundary. This fact appears more clearly in the second problem (down) where the set of classes is completely mixed and, therefore, with no linear separation. We underlined some areas of relevance over the feature space. For instance, the bottom left area is of particular interest. As we can see, the blue point appears to be an outlier in a region which has an higher density of red points. The 1-NN selection will only define the boundaries based on proximity. Oppositely, the HV-MOTS method clearly exhibits this outlier nature as it accounts for the *spread* and, consequently, the density of local points.

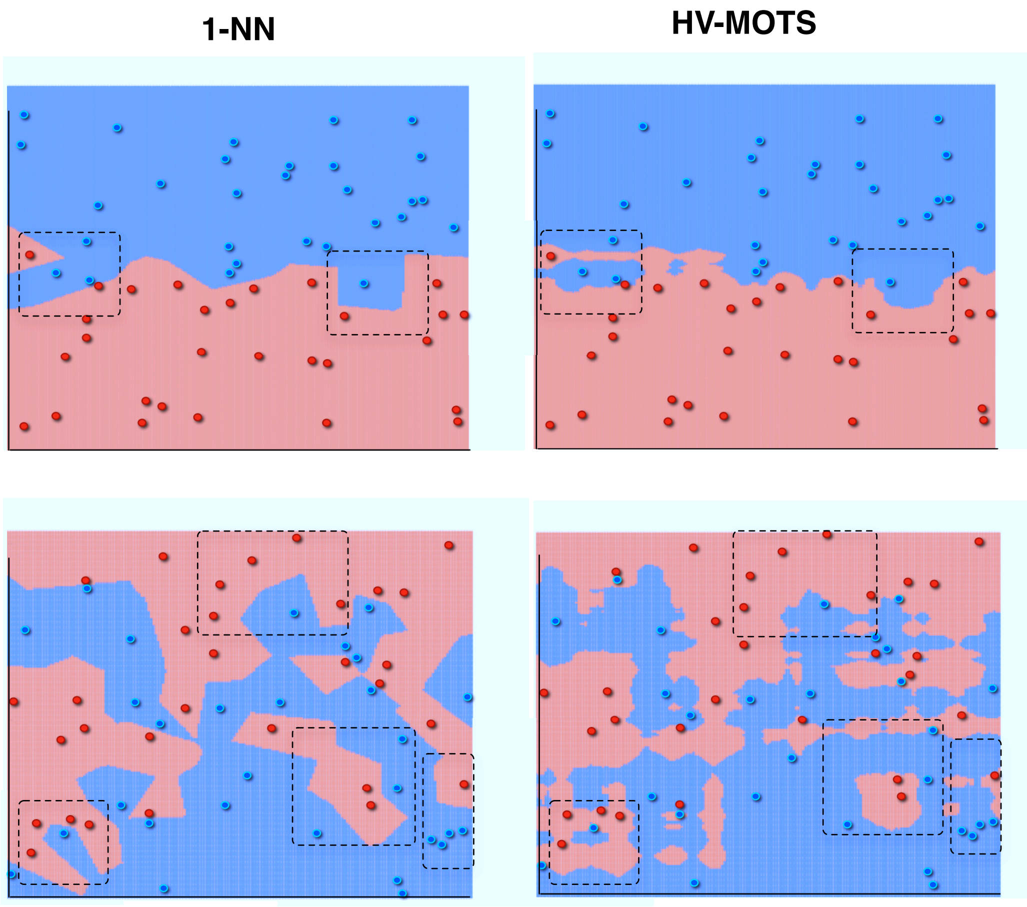



Figure 36: Comparison of the classification boundaries represented in feature space implied by 1-NN selection or HV-MOTS classification algorithms. The first problem (up) represents synthetic data where the classes are *almost* linearly separable. The second problem (down) represents a mixed set of classes data with no linear separation.

# 9 LARGE SCALE STUDY

In order to perform a thorough evaluation of the efficiency of the HV-MOTS classification framework, we follow several guidelines provided by past research in multiple classifiers comparison [102, 193, 252, 314, 315]. Therefore, we test our approach on a wide range of datasets that covers a variety of several scientific fields. This allows to assess extensively our proposal and at the same time provide conclusions that are free from any *data bias*. Moreover, we can also analyze in *which* situations the HV-MOTS classifier performs best. This entails the scientific field to which it is applied, but also the *characteristics* of the problem (number of classes, samples and objectives). Finally, this evaluation framework also offers an opportunity to analyze which features were chosen to obtain the best classification accuracy. The analysis of this choice can provide some insights on the underlying classification problem.

Therefore, we collected several datasets that all meet the same requirements. First, they should be focused on *time series* data and, therefore, on processes that are inherently temporal. Secondly, they should be structured around a *classification* task. In addition, they should also have a *multidimensional nature* and, therefore, provide at least two objectives or features. Finally, these datasets should optimally be linked to research papers that provide state-of-art classification results. Therefore, we gathered corresponding accuracies, characteristics and methods used for further comparison.

## 9.1 DATASETS SUMMARY

We provide in this section a brief summary of the datasets collected for the evaluation. For the sake of clarity, we only give an *overview* of the datasets here. However, for the sake of completeness, the complete description of datasets along with technical informations is available in the annex of this document (cf. Section A).

Overall, we collected 40 datasets, which characterize a range of classification problems. These datasets come from fields of *speech recognition*, *handwriting analysis*, *character recognition, movement analysis*, *hand sign recognition, brain-computer interface, EEG analysis, MEG analysis, ECoG analysis, cardiology, medical surveillance, climatology, radar analysis* and *robotics*. The characteristics of these datasets exhibit wide differences. The number of features varies between 2 and 274 (mean : 40.75, median : 12), the number of classes varies between 2 and 183 (mean : 20.9, median : 8) and the number of samples varies between 47 and 164860 (mean : 7540.12, median : 1347).

Regarding the features themselves, their properties also vary widely. This once again allows to abstract the potential conclusions from data bias. The corresponding time series span from 36 to 2134 time points which represent various underlying durations. The corresponding sizes of datasets range from 4.9 Mo to 1721 Mo (mean : 211.448 Mo, median : 83 Mo) for a complete size of 11572.04 Mo.

| Datasets | | Description | Feats | Class | Samples |
|---|---|---|---|---|---|
| Arabic digit | [158] | Spoken arabic digits | 13 | 10 | 8800 |
| Artificial characters | [52] | Character recognition | 2 | 10 | 6000 |
| Australian signs | [186] | Hand sign recognition | 10 | 95 | 6650 |
| Australian signs (HQ) | [187] | Hand signs (HQ) | 20 | 95 | 2565 |
| BciIII-03a-Graz | [48] | Brain-Computer | 60 | 4 | 840 |
| BciIV-01-Berlin | [49] | Brain-Computer | 64 | 3 | 1400 |
| BciIV-03-Freiburg | [49] | Brain-Computer | 10 | 4 | 480 |
| Biomag-2010 | [359] | EEG analysis | 274 | 2 | 780 |
| Challenge-2011 | [384] | Cardiology | 12 | 2 | 2000 |
| Character-trajectories | [276] | Character recognition | 3 | 20 | 2858 |
| Dachstein | [151] | High altitude medicine | 3 | 2 | 698 |
| Eeg-alcoholism | [408] | Medical analysis | 64 | 6 | 650 |
| Forte-2 | [41] | Climatology | 2 | 7 | 121 |
| Forte-6 | [41] | Climatology | 6 | 7 | 121 |
| Gaitpdb | [219] | Gait analysis | 18 | 2 | 306 |
| Handwritten | [126] | Character recognition | 2 | 183 | 8235 |
| Ionosphere | [114] | Radar analysis | 34 | 2 | 358 |
| Japanese-vowels | [213] | Speech analysis | 12 | 9 | 640 |
| Libras | [318] | Movement recognition | 2 | 15 | 360 |
| Pen-chars-35 | [286] | Character recognition | 2 | 62 | 1364 |
| Pen-chars-97 | [78] | Character recognition | 2 | 97 | 11640 |
| Person activity | [189] | Movement analysis | 12 | 11 | 164860 |
| Physical action | [353] | Movement analysis | 8 | 20 | 80 |
| Ptbdb-1 | [150] | Cardiology | 15 | 9 | 2750 |
| Ptbdb-2 | [150] | Cardiology | 15 | 9 | 2750 |
| Ptbdb-5 | [150] | Cardiology | 15 | 9 | 2750 |
| Robot failures-1 | [66] | Robotics | 6 | 4 | 88 |
| Robot failures-2 | [66] | Robotics | 6 | 5 | 47 |
| Robot failures-3 | [66] | Robotics | 6 | 4 | 47 |
| Robot failures-4 | [66] | Robotics | 6 | 3 | 117 |
| Robot failures-5 | [66] | Robotics | 6 | 5 | 164 |
| Slpdb | [61] | Sleep apnea analysis | 7 | 7 | 4085 |
| Sonar | [348] | Sonar analysis | 60 | 2 | 208 |
| Synemp | [184] | Climatology | 2 | 2 | 20000 |
| Vfdb | [212] | Cardiology | 2 | 15 | 600 |
| Vicon physical | [353] | Physiological analysis | 26 | 20 | 2000 |
| Wall-robot-4 | [127] | Robotics | 28 | 4 | 5460 |
| Wall-robot-24 | [127] | Robotics | 28 | 24 | 5460 |

## 9.2 METHODOLOGY

We intend to compare the classification results between the HV-MOTS framework and several classifiers. Therefore, we will outline in the next section the guidelines from previous research on the comparison of multiple classifiers and datasets. First, we underline the argument put forward by Keogh and Kasetty [194] , that such comparisons should try to be free of both *implementation* and *data bias*.

**Definition 32.** *Implementation bias* is the conscious or unconscious disparity in the quality of implementation between a proposed approach and the competing approaches.

As proposed by [194], we tried to perform extremely conscientious implementations of all approaches, combined with diligent explanations of the experimental process.

**Definition 33.** *Data bias* is the conscious or unconscious use of a particular set of testing data to confirm a desired finding.

In order to avoid data bias, we present the results from all the gathered datasets. These sets are from several scientific fields, with various properties, number of features, classes and samples in order to maximize the diversity of evaluation data.

### 9.2.1 *Evaluation framework*

In order to perform a comprehensive and thorough evaluation of the HV-MOTS classification scheme, we follow the guidelines provided by several researches focused on the comparison of classifiers over multiple datasets [102, 252, 314, 315]. That way, we try to avoid falling into the pitfalls of comparisons that could mislead the readers. As noted by [314], extremely large databases are bound to contains some statistical anomalies. Hence, careful attention should be directed to these phenomena that could invalidate experimental comparisons. Therefore, we avoid the *multiplicity effect* and do not use the simple paired t-test, but rather use the non-parametric Friedman test, as also proposed by [102], with the corresponding post-hoc tests for comparing all classifiers. Therefore, we first use Tukey-Kramer Honestly Significant Difference (HSD) test [94] over the results of Friedman's ANOVA, as proposed by [102] and also suggested in other fields of evaluation [110]. Finally, we present the *critical difference graphs* [102] which allows to exhibit the true statistical superiority and eventual groups of statistical equivalence between various methods. As pointed by [314], the *repeated careful tuning* is usually performed in order to produce a desired result on a particular set of data. Hence, the results produced by this methodology might be misleading. Therefore, *we do not operate any form of tuning* nor do we make *any kind of pre-processing* on the datasets in order to enhance results. We underline the fact that we do not perform any further extraction of features and work solely on the raw time series available. That way, we focus solely on the classification method used rather than the discriminative power of the underlying feature sets.

In order to compute the similarity between elements, we directly use the time series features in each dataset. We first normalize the series using *zero mean* and *unit variance* transformations. The mean and deviation of the time series are kept and used separately as features. Finally, to compare the time series, we use the Dynamic Time Warping (DTW) distance and the Euclidean distance computed on down-sampled series, which leads to two different distance measures per time series. Therefore, we have four different features comparisons (mean, deviation, DTW and Euclidean) for each time series. This leads to a distance matrix computed between every element of the dataset.

As our method does not require any training, we use the *Leave-One-Out* evaluation methodology. That is, each file is first withdrawn from the dataset and then input for classification with the remaining set acting as a database. In order to compare different methods, we perform large-scale experiments by testing combinatorial possibilities among every available feature for each dataset. Therefore, we start by performing classification with only one feature. We then test classification with every combination of two features, and so forth. Given that this testing methodology implies an exponentially growing number of tests, we keep only the top performing half of the features set after each step. The selection of retained features is based on their mean classification accuracies (across methods). We repeat this procedure and halve the set of available features (in which to choose the objectives for classification) until the number of remaining features is less than the current number of objectives. This testing methodology ensure *completeness* as we will test the statistical significance independently of the underlying feature selection. Therefore, we test the classification power of each method on *every* subset of available features. Hence, we focus on the *classification criterion* instead of the set of features used. Furthermore, this methodology also allows to perform a *feature analysis*, by extracting for each dataset the best performing features combinations.

### 9.2.2 *Hardware*

As our testing methodology is computationally intensive, it required a high-performance supercomputing ressource in order to attain completion. We made all our calculations on the Guillimin cluster from the CLUMEQ supercomputing center of McGill university under the govern of Calcul Quebec and Calcul Canada. The exploitation of this supercomputer is financed by the Canadian fondation for innovation (FCI), the Research Concil of natural sciences (CRSNG), NanoQuebec, the RMGA and the Quebec research fund - Nature and technologies (FRQ-NT). The Guillimin server contains 1200 computing nodes and 34 infrastructure nodes. These nodes are all constituted by a pair of Intel Westmere-EP (Xeon X5650) processors each of which containing 6 processing cores and 24, 36 ou 72 gigabytes of RAM memory. All nodes are linked together by a high performance Infiniband QDR network.

### 9.2.3 *Algorithms implementation*

In order to perform an exhaustive evaluation, we tried to implement several classification techniques. However, it should be noted that we are working with *time series* data which implies a different set of constraints than the usual classification problems. In time series classification, the best published results are usually provided by simple nearest neighbor methods [382]. As discussed earlier, the 1-NN method has been proven as the most efficient classifier over a wide variety of datasets [194]. Nevertheless, we also implemented the 5-NN, NC and SVM classifiers to ensure a thorough comparison. The 1-NN, 5-NN and NC classification techniques can be directly applied to the distance matrix. However, the SVM classifier usually requires a matrix of features for classification rather than a matrix of distances. Furthermore, we must face the high dimensionality of the datasets and can not use the time series features as direct input to an SVM. Therefore, we follow the proposal of Gudmundsson et al. [149] and use a *proximity function* kernel which is designed to classify pairwise data.

1-NN — We find the nearest element to the input (based on the norm of the distances) by computing the complete distance matrix and selecting the class accordingly.

5-NN — The same idea applies to finding the five nearest element of the dataset and then selecting the corresponding class.

NC — The centroid of each class is computed based on the distance matrix. We then select the class with the nearest centroid.

SVM — As proposed by Gudmundsson et al. [149] we implement a SVM based on the *pairwise-proximity function* kernel (ppfSVM), which allows to use directly the distance matrix in order to perform classification.

### 9.2.4 *Reproducibility of experiments*

In order to allow interested readers to reproduce our experiments but also for further research and comparison with other classification techniques, we made all the datasets available on a dedicated web page[1]. The complete source code of the testbed is also available. However, using each dataset requires to report proper credits (the different references are provided in the annex of this document).

## 9.3 RESULTS AND ANALYSIS

We present in this section the results of the large scale study. We start by giving a detailed dataset-wise view on the results (Section 9.3.1). We exhibit in this analysis the statistical superiority of the HV-MOTS classifier over other schemes. Then, we try to obtain a higher-level view on the results by analyzing the potential influence of the number of features, samples, classes and domains on these statistical differences (Section 9.3.2).

### 9.3.1 *Dataset-wise results*

*Overall accuracy*

We start by providing the complete classification results separated over each dataset for the *1-NN*, *5-NN*, *NC*, *SVM*, *MOTS*, *NP-MOTS* and *HV-MOTS* classifiers. We present in Table 3 the best classification accuracies achieved for each method.

The dark grey rows indicate the datasets in which the HV-MOTS classifier outperforms other methods. The light grey rows are datasets where it is tied in first position. As we can see, the HV-MOTS approach ranks first in 33 out of 40 datasets with a tied best result in 4 of these datasets. As indicated by the *no free lunch* theorem [381] there is not a *single* dominant classifier for *all* classification problems. However, the HV-MOTS approach appears to be superior in the majority of cases. Furthermore, even for the few datasets in which it is dominated, the HV-MOTS classifier remains a competitive approach.

As we discussed earlier, the reported classification accuracies are insufficient to draw solid conclusion when comparing classifiers [102, 314]. In order to confirm the previous analysis of results through statistical significance, we present in Figure 37 the results

1 http://repmus.ircam.fr/esling/hvmots-datasets.html

| | 1-NN | 5-NN | NC | SVM | MOTS | NP-M. | HV-M. |
|---|---|---|---|---|---|---|---|
| Arabic digit | 99.54 | 99.26 | 83.37 | 99.31 | 99.61 | 99.52 | **99.88** |
| Artificial characters | 99.98 | 99.98 | 96.40 | 99.98 | **100.0** | 99.98 | **100.0** |
| Australian signs | 71.52 | 70.23 | 65.85 | 72.33 | 72.35 | 69.23 | **84.36** |
| Australian signs (HQ) | 84.64 | 82.46 | 70.25 | 85.22 | 64.44 | 76.14 | **91.31** |
| BciIII-03a-Graz | 34.17 | 33.21 | 25.12 | 33.81 | 33.93 | 32.62 | **35.59** |
| BciIV-01-Berlin | 57.29 | 56.93 | 52.14 | 56.54 | 56.92 | 56.43 | **57.50** |
| BciIV-03-Freiburg | 33.40 | 33.19 | 32.76 | 33.40 | 36.40 | 34.04 | **36.62** |
| Biomag-2010 | 67.12 | 68.36 | 71.12 | **73.42** | 72.24 | 69.62 | **73.42** |
| Challenge-2011 | 90.58 | 92.19 | 88.38 | **92.69** | 90.58 | 90.08 | 91.98 |
| Character-trajectories | **99.23** | 98.67 | 89.68 | 99.22 | 98.11 | 99.16 | **99.23** |
| Dachstein | 97.42 | 97.13 | 89.67 | 98.28 | 98.13 | 98.13 | **98.57** |
| Eeg-alcoholism | 79.55 | 81.36 | 80.00 | 80.06 | 82.27 | 79.55 | **83.18** |
| Forte-2 | 82.64 | 80.17 | 74.38 | 82.45 | 81.54 | 82.64 | **83.47** |
| Forte-6 | 71.90 | 70.25 | 62.81 | 71.56 | 71.90 | 71.90 | **73.55** |
| Gaitpdb | **89.87** | 83.01 | 73.20 | 87.24 | 77.78 | 76.14 | **89.87** |
| Handwritten | 90.67 | 82.34 | 78.12 | 84.72 | 81.23 | 80.32 | **92.17** |
| Ionosphere | 94.20 | 92.31 | 88.89 | 93.12 | 89.17 | 90.88 | **94.59** |
| Japanese-vowels | 94.21 | 95.00 | 88.91 | 94.77 | 90.78 | 92.03 | **97.19** |
| Libras | 84.44 | 77.78 | 55.00 | 83.79 | 79.72 | 75.56 | **91.39** |
| Pen-chars-35 | 54.32 | 52.18 | 27.94 | 54.12 | 51.44 | 49.59 | **54.91** |
| Pen-chars-97 | 47.69 | 45.95 | 20.66 | 46.80 | 50.52 | 47.69 | **56.80** |
| Person activity | 89.35 | 88.72 | 71.80 | 89.31 | 85.09 | 80.95 | **91.60** |
| Physical action | **95.00** | 92.50 | 90.00 | 94.50 | 92.50 | 92.50 | 93.75 |
| Psychophysics | 75.32 | 77.27 | 70.13 | 75.03 | 72.08 | 74.03 | **80.52** |
| Ptbdb-1 | 98.77 | 96.73 | 52.73 | 98.66 | 97.34 | 96.84 | **99.12** |
| Ptbdb-2 | 96.19 | 93.23 | 46.50 | 95.29 | 92.73 | 91.12 | **97.58** |
| Ptbdb-5 | 91.19 | 87.54 | 42.15 | 90.00 | 87.23 | 83.31 | **93.73** |
| Robot failures-1 | 96.59 | 96.59 | 76.13 | 96.16 | 92.04 | 89.77 | **97.73** |
| Robot failures-2 | 78.72 | 76.59 | 72.34 | 76.49 | 76.59 | **82.97** | 78.72 |
| Robot failures-3 | 86.97 | 86.17 | 71.17 | 86.36 | 85.83 | 83.99 | **91.49** |
| Robot failures-4 | 94.87 | 95.72 | 74.36 | 95.26 | **98.29** | 88.03 | **98.29** |
| Robot failures-5 | 77.44 | 76.89 | 62.80 | 76.93 | 75.00 | 71.95 | **79.88** |
| Slpdb | 77.86 | **81.18** | 72.78 | 79.81 | 79.09 | 74.71 | 79.56 |
| Sonar | 86.54 | 85.10 | 70.67 | 86.35 | 86.54 | 87.02 | **88.94** |
| Synemp | 91.15 | 87.35 | 80.82 | 90.24 | 85.28 | 76.23 | **92.28** |
| Vfdb | 60.64 | **63.18** | 43.58 | 63.05 | 56.08 | 48.47 | 58.62 |
| Vicon physical | **97.00** | 95.50 | 89.00 | 96.50 | 94.00 | 94.00 | 96.00 |
| Wall-robot-4 | 96.55 | 96.55 | 50.02 | 97.31 | 97.31 | 97.31 | **97.36** |
| Wall-robot-24 | 91.92 | 90.76 | 41.60 | 91.23 | 91.46 | 91.42 | **92.47** |

Table 3: Comparison of overall classification accuracies for different methods

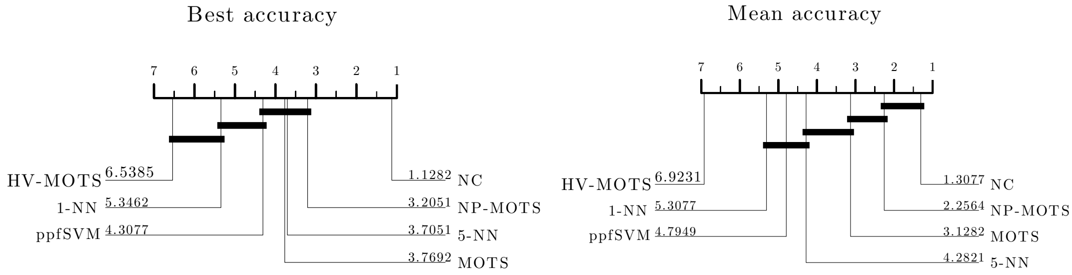



Figure 37: Critical difference graphs for the best (left) and mean (right) accuracies over every dataset.

of a post-hoc Nemenyi test. These results are shown as *critical difference graphs* for *best* (left) and *mean* (right) accuracies. Hence, the figure on the left is the direct application of the results from Table 3. The figure on the right is computed by using the *mean* classification accuracies for each dataset over *every* features combination.

In both cases, the HV-MOTS classifier ranks first in statistical rank differences. In the best accuracy results, the HV-MOTS approach belongs to the same *clique* than the 1-NN classifier. This indicates that even if HV-MOTS almost always rank above 1-NN, their classification results are in the same range of accuracy. If we look back at Table 3, we can see that the 1-NN classifier almost always ranks second. Therefore, it outperforms remaining classifiers, which correlates with previous findings in time series classification. The same observations hold for the 1-NN and ppfSVM which appears to belong to the same classification ranges. Finally, the ppfSVM, MOTS, 5-NN and NP-MOTS belong to the same clique and the NC cis statistically inferior to all the classification approaches. If we look at the *mean* results, it seems that HV-MOTS exhibits a critical difference with the other approaches. This indicates that given *any* set of features, the HV-MOTS is statistically superior. Hence, it seems that the introduction of a multidimensional assessment of similarity does not only benefit audio problematics but can also improve a wide variety of research fields. We have willingly left the audio classification problems aside in order to see if the concepts proposed in our approach could benefit *other* research topics. However, Section 11 will be entirely dedicated to the application of the HV-MOTS approach to audio classification.

*Statistical significance*

The overall analysis of results allowed to exhibit the statistical superiority of the HV-MOTS approach. In order to provide an in-depth analysis of these results, we present in Figure 38 the statistical significance tests for each dataset separately. In this figure, the mean column ranks are obtained from Tukey-Kramer HSD based on a Friedman's ANOVA [102]. These statistical tests are performed on the accuracy matrix of *every* features combination available for each dataset. Hence, we can see which method is significantly superior, independently of the features used for classification. This allows to focus on the classification *criterion* rather than the *feature selection*. We add to these graphs the range of statistical significance for the HV-MOTS ranks (dotted horizontal lines).

As we can see, the HV-MOTS classifier is statistically superior for almost every dataset. We can also see that, for some of the datasets in which the HV-MOTS approach does not provide the *best* classification (*physical-action*, *vfdb* and *vicon-physical*), it still

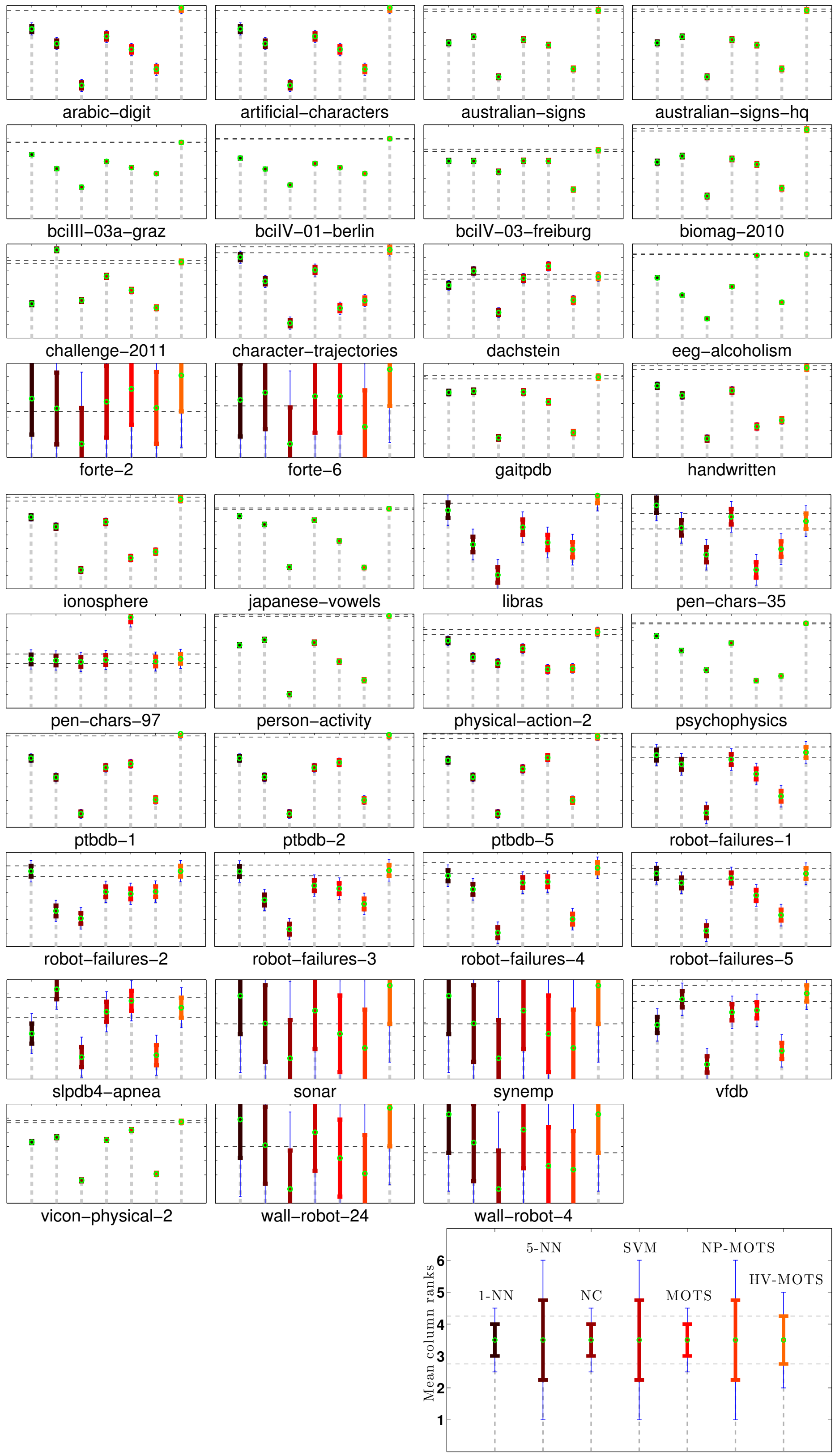



Figure 38: Comparison of statistical significance between classification methods based on the Tukey-Kramer HSD over Friedman's ANOVA.

obtains the best statistical ranks. This means that even though it does not provide the best accuracy score, the HV-MOTS classifier will still rank first for most of the features combinations. The reciprocal observation holds for the *dachstein* and *penchars* datasets, where it seems that the 1-NN and MOTS approaches obtain the best mean ranks. Finally, some of the datasets (*forte, sonar, synemp* and *wall-robot*) provides more ambiguous results. These results seem to indicate that, in the corresponding datasets, the accuracy results will vary more widely depending on the various features combinations.

### 9.3.2 *Global scale analysis*

In this section, we try to extract some higher-level clues on the statistical differences. We examine the different properties of the datasets, in order to see if the previous results can be explained by variations in these characteristics. Hence, we try to extract the eventual relationships between the number of *features*, *classes* and *samples* and the classification results. In order to perform this analysis, all the following results are computed with a Tukey-Kramer HSD over a Friedman ANOVA. Whenever we analyze a different aspect of variability, we concatenate all datasets that fall under a particular category and then compute the entire ANOVA. For example, in the *class-wise* analysis, the results of all 2-class datasets are concatenated in order to obtain a complete data matrix to analyze. Then, the same procedure is applied with the higher numbers of classes. We perform this analysis for *class* cardinality, *samples* cardinality, *features* cardinality and finally regroup results depending on the scientific *domain* of study. It should be underlined that these comparisons are performed only with the datasets of the present study. Hence, some of the groupings might lack sufficient number of examples to provide a clear correlation in the results.

#### *Classes cardinality analysis*

First, we analyze the statistical differences between methods depending on the number of classes in the datasets. Therefore, we regroup classification results from datasets that share the same number of classes. We perform the statistical significance analysis on each accuracy matrix separately. The statistical differences are presented in Figure 39.

As we can see, the HV-MOTS classifier is statistically superior for every cardinality. It seems that, for an increasing number of classes, this superiority increases as well. However, the statistical difference is not *significant* for two of the low class cardinalities. In the *2 classes* dataset, the 5-NN classifier appears to be in the same statistical range as HV-MOTS. For the *5+* class, both the 1-NN and MOTS approaches are in the same statistical range. Finally, the largest statistical differences are provided by the highest number of classes (*10+, 20+* and *50+*). For these datasets, the performances of other classifiers seems to follow the same relative rankings. Hence, it appears that the HV-MOTS classifier is robust for higher numbers of classes in the dataset. Regarding the accuracy of other classifiers, it seems that the 1-NN method obtain an equivalent ranking and is usually in the same range as the SVM and 5-NN classifiers. The MOTS approach appears to suffer the most for cardinalities higher than 5 classes in the dataset. This can be correlated to the nature of its selection method, which is based on counting the number of occurrences of classes. Hence, it appears logical that a higher number of classes will increase the probability of confusions when counting the occurences in a Pareto front. Finally, it is interesting to note the "co-evolution" of the results for NC and NP-MOTS classifiers. As both methods are based on computing the mean

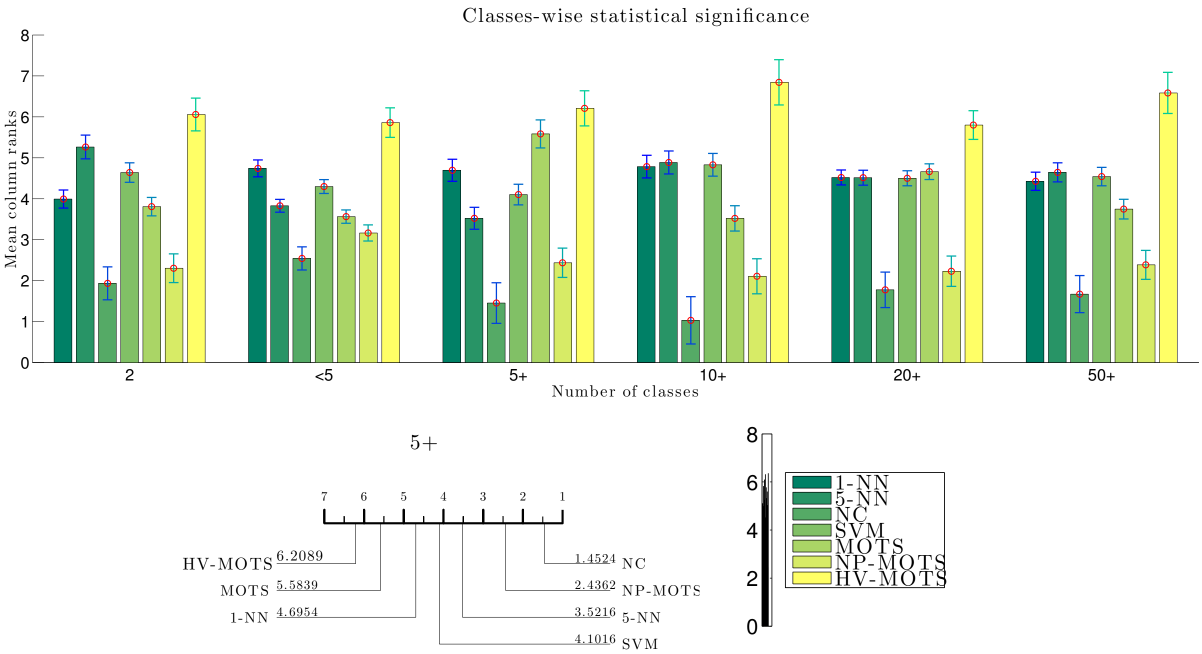



Figure 39: Comparison of statistical significance between classification methods for an increasing number of classes based on the Tukey-Kramer HSD over Friedman's ANOVA. The *critical difference graph* is presented for the 5+ classes.

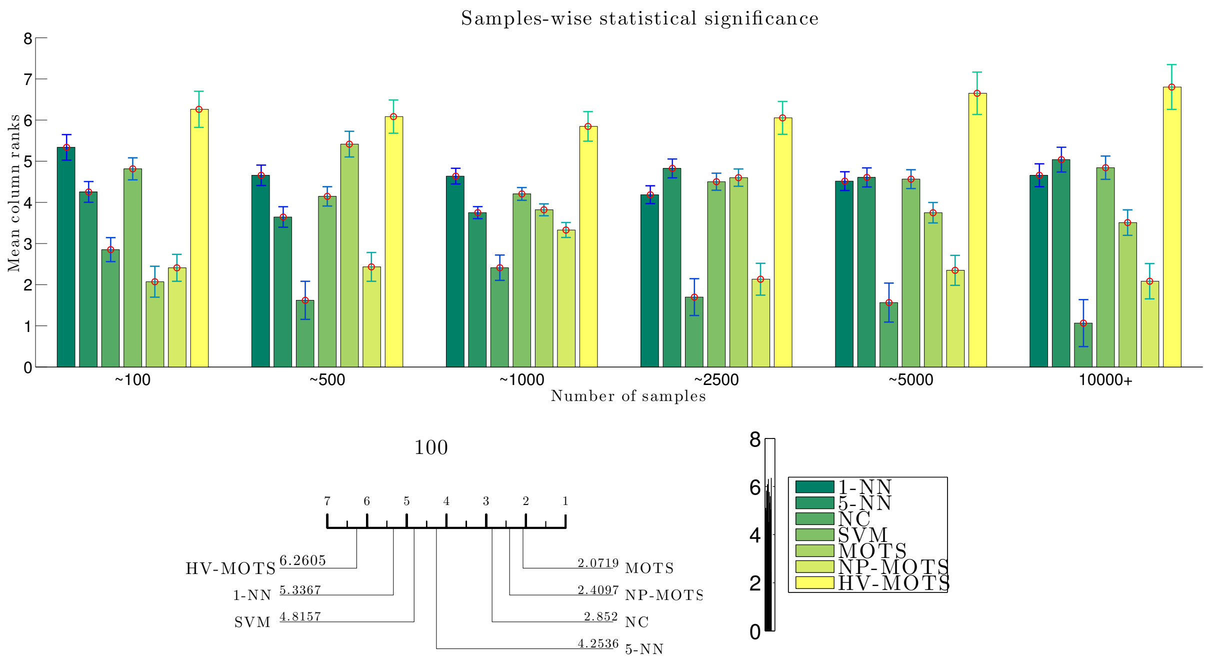



Figure 40: Comparison of statistical significance between classification methods for an increasing number of samples based on the Tukey-Kramer HSD over Friedman's ANOVA

distances between elements of the classes, they appear to provide equivalent results. Indeed, we can interpret the NP-MOTS classifier as a NC selection applied solely to the Pareto front of classes.

*Samples cardinality analysis*

We analyze the statistical differences between methods depending on the number of samples in the dataset. Figure 40 shows the results this analysis.

As we can see, the HV-MOTS classifier appears to increase in statistical superiority with a higher number of samples to classify. Hence, it seems that the HV-MOTS classification benefits from a larger number of instances in each class. This can be correlated to the computation of the hypervolume itself. Indeed, as the HV-MOTS selection studies the behavior of the whole class, a higher number of points in each class might provide a finer approximation of its true hypervolume. This can be best illustrated by looking back at Figure 33. We can see that the computed hypervolume is a quite coarse approximation. Hence, if we suppose that some supplementary points interleave the existing ones, we would get a much finer estimation of the hypervolume. In order to extend this result to the performance of other classifiers, it is interesting to note the evolution of 1-NN and 5-NN methods. It seems that even if the 1-NN selection performs higher for lower samples cardinalities, the performances of the 5-NN method increase with a higher number of samples. This correlates with the

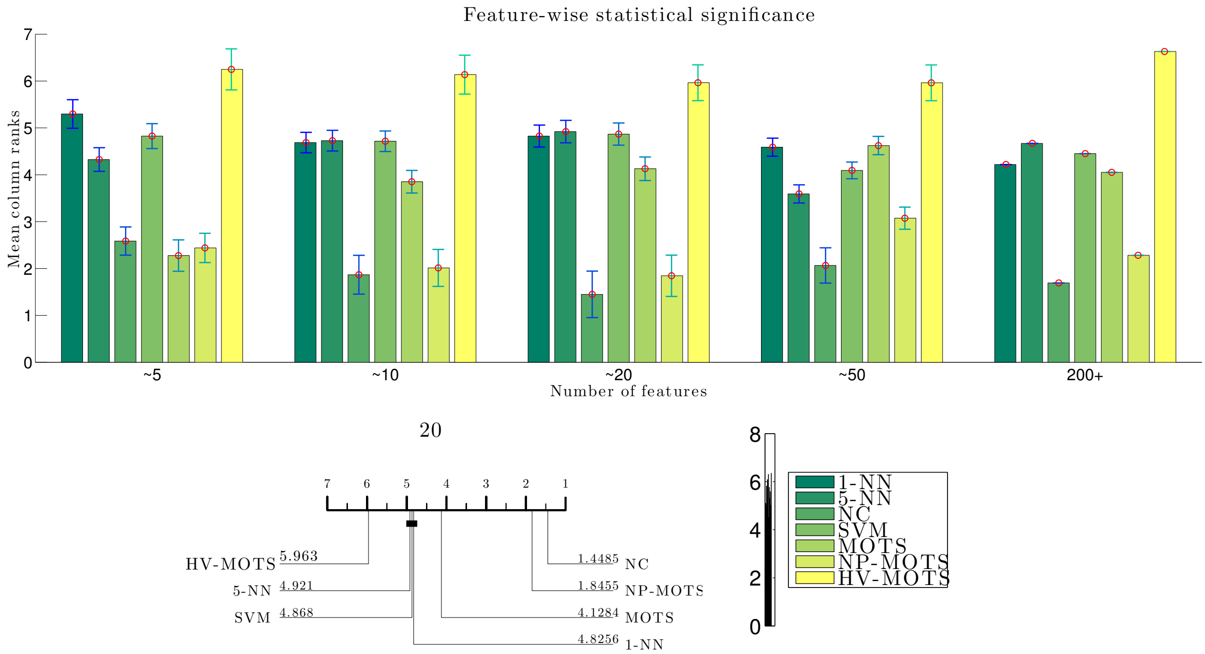



Figure 41: Comparison of statistical significance between classification methods for an increasing number of features based on the Tukey-Kramer HSD over Friedman's ANOVA

previous observation, as both HV-MOTS and 5-NN study the behavior of multiple instances from the classes, they benefit from a higher number of samples. Finally, it appears that the performances of the NC and NP-MOTS classification still denote a joint evolution.

*Features cardinality analysis*

We analyze the statistical differences between methods depending on the number of features in the dataset. Figure 40 shows the results of this analysis.

As we can see, the performances of HV-MOTS appear to be steadily superior to other classification schemes. However, it is interesting to focus on the evolution of *all other* methods. Indeed, it seems that other classifiers are suffering from a joint decrease in their statistical ranks. Hence, the statistical *differences* increase with a higher number of features. We believe that this could be explained, not by the properties of the datasets nor classification methods but by the evaluation methodology itself. Indeed, we tested all combinatorial possibilities inside the sets of available features. Furthermore, we have shown in Figure 37 that the HV-MOTS classifier is significantly superior in *mean* classification accuracy. Hence, by linking these two observations together, we can hypothesize that there might be a *multiplicity effect* that appears with a higher number of features. This also seems confirmed by the variances of statistical ranks, which

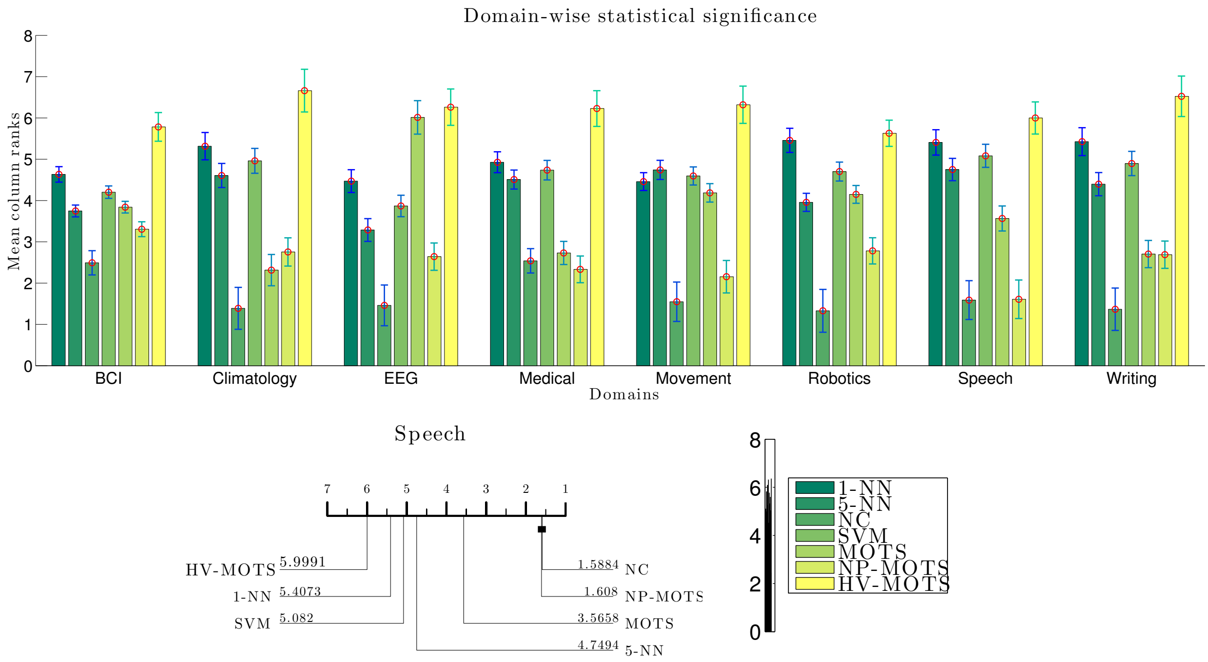



Figure 42: Comparison of statistical significance between classification methods depending on the scientific domain being studied based on the Tukey-Kramer HSD over Friedman's ANOVA

exhibit a steady decrease. Therefore, we avoid drawing some conclusions from the features cardinality analysis as they might be influenced by this effect.

*Domain-wise analysis*

Finally, we perform a domain-specific analysis in order to regroup dataset depending on their topics. Figure 42 summarizes these results.

As we can see, the performances of classifiers do not seem to depend highly on the nature of the data. Even though the HV-MOTS classification appears to be generally superior, some of the domains does not provide a *significant* superiority. This is the case for *robotics*, *speech* and *EEG* analysis where other classifiers are in the same statistical range. However, a true statistical significance appears for *BCI, climatology, movement, writing* and *medical.* Finally, the overall results in medical datasets also motivate our next chapter on heart sounds biometry.

## 9.4 COMPARISON TO STATE-OF-ART RESULTS

We provide a brief comparison between the HV-MOTS results and state-of-art accuracies and methods. This comparison is based on the classification accuracy reported in each corresponding paper. We also perform this comparison with the 1-NN best classification accuracy. Table 4 summarizes the comparison to state-of-art algorithms on each dataset. We also provide in this table the reference classification methods used and the corresponding reported accuracy.

The first striking observation is the wealth and diversity of classification methods that have been used for these datasets. Furthermore, each of these researches have applied different pre-processing steps to the features. Moreover, the corresponding methods usually undergo a particularly careful and extensive parameter tuning. We underline the fact that we did not perform any kind of features pre-processing, transformation or modification. Therefore, these results are based solely on the available raw time series and corresponding mean and deviation features. Furthermore, as the HV-MOTS approach does not require any settings, we did not perform any parameter tuning. The darker gray rows indicate the datasets for which the HV-MOTS classifier obtains better results than the published state-of-art accuracy. The lighter gray rows show the datasets where HV-MOTS performs best, but because of differences in the evaluation methodologies, the results can not be compared straightforwardly. As we can see, improved performances are provided for 28 datasets. For the remaining datasets, the HV-MOTS approach is usually competitive, even without advanced data transformation and parameters tuning. However, if we look specifically at the BCI datasets, it appears that the results we obtained are extremely inferior to those reported in the literature. This may indicate that these datasets require a mandatory feature transformation step. Indeed, in this case, the relevant information is strongly diluted inside the wealth of data provided by the array of EEG electrodes.

## 9.5 Extended analysis

In this section, we provide an extended analysis of more specific aspects of the results. First, we present a detailed analysis of the features selected to obtain the best classification results in some of the datasets. This allows to get a deeper understanding on the reasons *why* the classification is accurate. Then, we show the influence of using the time series data rather than mean and deviation statistics. Finally, we provide an analysis on the influence of different parameters over the time series matching.

### 9.5.1 *Selected features*

The main advantage of our evaluation methodology is that it gives us access to the individual accuracy of every feature available for each dataset. Therefore, we can easily investigate the set of features automatically selected by the testbed to obtain the best classification accuracy. We focus our attention here to a part of the datasets collection that provides some compelling observations based on their selected features.

First, as we can see in the *australian signs* dataset, the best features selected are a combination of information coming from several sources. Therefore, to perform the best classification, a set of complementary coordinates is selected. In this combination, all the features are time series information. Then, the multidimensional assessment of similarity allows to study these dimensions jointly. It is interesting to note that it is an alternation between the left and right coordinates for the same parts of the hands that allow to discriminate between different signs. The same observations can be made for the *gaitpdb* dataset. Indeed, a joint assessment of the left and right evolution of the body allows to provide the best detection of the instability caused by Parkinson's disease. Finally, for the *person activity* dataset, it is the simultaneous analysis of the

| Name | Algorithm | Results | 1-NN | HV-MOTS |
|---|---|---|---|---|
| Arabic digit | Vector Quantization + Tree | 93.12 | 99.54 | **99.88** |
| Artificial characters | Genetic Algorithm | 98.68 | 99.98 | **100** |
| Australian signs | Hidden Markov Model | 71.20 | 71.52 | **84.36** |
| Australian signs HQ | Tree Class Algorithm | **94.50** | 84.64 | 91.31 |
| BciIII-03a-Graz | Multi-Class CSP - Fisher | **79.26** | 34.17 | 35.59 |
| BciIV-01-Berlin | Principal Component Analysis | **62.80** | 57.29 | 57.50 |
| BciIV-03-Freiburg | SVM + LDA | **46.90** | 33.40 | 36.62 |
| Biomag-2010 | SVM | 69.00 | 67.12 | **73.42** |
| Challenge-2011 | Matrix of regularity | 85.90 | 90.58 | **91.98** |
| Character-trajectory | HMM + GMM | 93.67 | **99.23** | **99.23** |
| Dachstein | - | - | 97.42 | **98.57** |
| Eeg-alcoholism | Multivariate HMM | 78.50 | 79.55 | **83.18** |
| Forte-2 | - | - | 82.64 | **83.47** |
| Forte-6 | Shared-NN | 60.84 | 71.90 | **73.55** |
| Gaitpdb | Neural Network - Wavelets | 77.33 | **89.87** | **89.87** |
| Handwritten | HMM + SVM | 92.00 | 90.67 | **92.17** |
| Ionosphere | Genetic Programming | 94.20 | 94.20 | **94.59** |
| Japanese-vowels | 5-state continuous HMM | 96.20 | 94.21 | **97.19** |
| Libras | Spiking Neural Network | 88.59 | 84.44 | **91.39** |
| Pen-chars-35 | DTW + NN | 89.15 | 54.32 | 54.91 |
| Pen-chars-97 | Template matching + NN | 91.80 | 47.69 | 56.80 |
| Person activity | Meta-Prediction Agents | 91.33 | 89.35 | **91.60** |
| Physical action | Genetic Programming | 73.30 | **95.00** | 93.75 |
| Ptbdb-1 | Random Forest | - | 98.77 | **99.12** |
| Ptbdb-2 | Random Forest | - | 96.19 | **97.58** |
| Ptbdb-5 | Random Forest | 85.80 | 91.19 | **93.73** |
| Robot failures-1 | Feature Transform + NN | 80.00 | 96.59 | **97.73** |
| Robot failures-2 | Feature Transform + NN | - | **78.72** | **78.72** |
| Robot failures-3 | Feature Transform + NN | - | 86.97 | **91.49** |
| Robot failures-4 | Feature Transform + NN | - | 93.87 | **98.29** |
| Robot failures-5 | Feature Transform + NN | - | 77.44 | **79.88** |
| Slpdb | Multi-Scale SVM | **88.97** | 77.86 | 79.56 |
| Sonar | Minimum Message Tree | 76.00 | 86.54 | **88.94** |
| Synemp | - | - | 91.15 | **92.28** |
| Vfdb | Filter + Peak detection | 91.50 | 60.64 | 58.62 |
| Vicon physical | Dynamic Neural Network | 95.40 | **97.00** | 96.00 |
| Wall-robot | Polynomial SVM | 95.58 | 96.55 | **97.36** |
| Wall-robot | Polynomial SVM | - | 91.92 | **92.47** |

Table 4: Comparison of classification accuracies with state-of-art results on the same datasets. We provide for each dataset the original algorithm used to obtained the reported classification accuracy.

| # | % | **Features** |
|---|---|---|
| **Australian-signs (HQ)** | | |
| 9 | [91.31] | $X^{l}_{64} - X^{r}_{64} - Roll^{l}_{128} - Roll^{r}_{128} - Yaw^{r}_{64} - Yaw^{r}_{64} - Pitch^{r}_{128} - Index^{r}_{128}$ |
| **Character-trajectories** | | |
| 4 | [99.23] | $X_{32} - Y_{32} - X_{\sigma} - Y_{\sigma}$ |
| **Dachstein** | | |
| 6 | [98.57] | $EEG^{C4}_{\sigma} - EEG^{C3}_{\sigma} - EEG^{C3}_{\mu} - ECG_{16} - ECG_{W.2\%} - ECG_{\sigma}$ |
| **Eeg-alcoholism** | | |
| 3 | [83.19] | $FP^{1}_{64} - CP^{1}_{W.2\%} - P^{6}_{\sigma}$ |
| **Gaitpdb** | | |
| 7 | [94.59] | $Left^{1}_{\mu} - Left^{1}_{\sigma} - Left^{4}_{16} - Right^{1}_{\mu} - Right^{2}_{\sigma} - Left^{Total}_{16} - Right^{Total}_{16}$ |
| **Japanese-vowels** | | |
| 6 | [97.19] | $LPC^{1}_{128} - LPC^{2}_{64} - LPC^{3}_{64} - LPC^{4}_{128} - LPC^{8}_{128} - LPC^{9}_{32}$ |
| **Libras** | | |
| 5 | [91.39] | $xCurve_{16} - xCurve_{\sigma} - yCurve_{16} - yCurve_{\sigma}$ |
| **Person activity** | | |
| 5 | [91.60] | $Ankle^{right}_{128} - Chest^{y}_{64} - Chest^{z}_{128} - Chest^{x}_{32} - Ankle^{left}_{128}$ |
| **Slpdb** | | |
| 4 | [79.56] | $BP_{16} - EEG_{\sigma} - Resp_{\mu} - Resp_{\sigma}$ |
| **Wall-robot** | | |
| 3 | [92.47] | $Sonar_{W.2\%} - Sonar_{\mu} - Sonar_{\sigma}$ |

Table 5: Comparison of classification accuracies with state-of-art results on the same datasets. We provide for each dataset the original algorithm used to obtained the reported classification accuracy.

left and right ankles, combined with the three-dimensional coordinates of the chest that allow to discriminate the different movements performed. Analogous observations can be made on the *character-trajectories* and *libras* datasets. However, it is interesting to note that, in these cases, it is the joint analysis of different coordinates of the *same* information that allows to perform the best discrimination.

If we turn our attention to the medical datasets, other interesting observations appear. First, in the *slpdb* dataset, the combination selected allows to get a sense of the data that are being studied. Indeed, this dataset focus on the detection of sleep apnea episodes. Hence, the selected features are the *blood pressure*, *variance* of the *EEG* and *mean* and *variance* of the *Respiration*. This seems remarkably to follow the intuition of the required features for assessing sleep apnea. In the same line of thought, the *dachstein* dataset gives information about the changes in human physiology at high altitude, with impacts on both *EEG* and *ECG* information. The results of the *eeg-alcoholism* dataset are also interesting in the fact that some of the electrodes from the EEG are selected rather than others. Therefore, this selection could give some clues on the impact of alcohol on the brain signals for future researches.

Finally, we selected the *Japanese-vowels* results. Indeed, they show that, inside a set of similar information (here the LPC bands), a multidimensional assessment of similarity can still provide an improvement of the classification results. To go even further in that observation, it is interesting to note that, for the *wall-robot* dataset, the three features come from the same source of information. However, this information is analyzed through its temporal evolution, standard deviation and mean values. Hence, this points to the fact that these might be complementary sets of information. In the following section, we will see that this observation appears to be confirmed throughout all of the datasets.

### 9.5.2 *The power of time*

Our approach was constructed on the assumption that the complete temporal information allows a better recognition and more accurate classification. In order to validate this hypothesis, we present in Figure 43 the dataset-wise analysis of results depending on the sets of information used. Therefore, this figure contains the classification results without time series features (*static* sets of information), with time series only (*temporal* sets without any mean or deviation information) and a combination of both sources (*mixed* sets).

As we can see, these results appear to confirm our hypotheses on the temporal information. Indeed, the use of time series data almost always improves the accuracy results of static information in both *mean* and *best* accuracy. Only the *gaitpdb*, *physical-action*, *robot-failures* and *vicon-physical* datasets do not follow this trend. For the two physical datasets, this can be explained by the nature of classification that is being performed. Indeed, these datasets try to discriminate between *aggressive* and *normal* action. If we look closely to the classification results, it appears that the *variance* (amplitude) of movements provides a better classification. It seems logical as in these cases, an *aggressive* action is the same as a *normal* action in terms of their temporal evolution, but it is its speed and amplitude that change its nature. The same kind of reasoning can be applied to the two other datasets. More interestingly, it appears that, for every dataset, the best results are always obtained by *mixed* sets of information. This result indicates that normalized temporal shapes and static values are complementary sets of information.

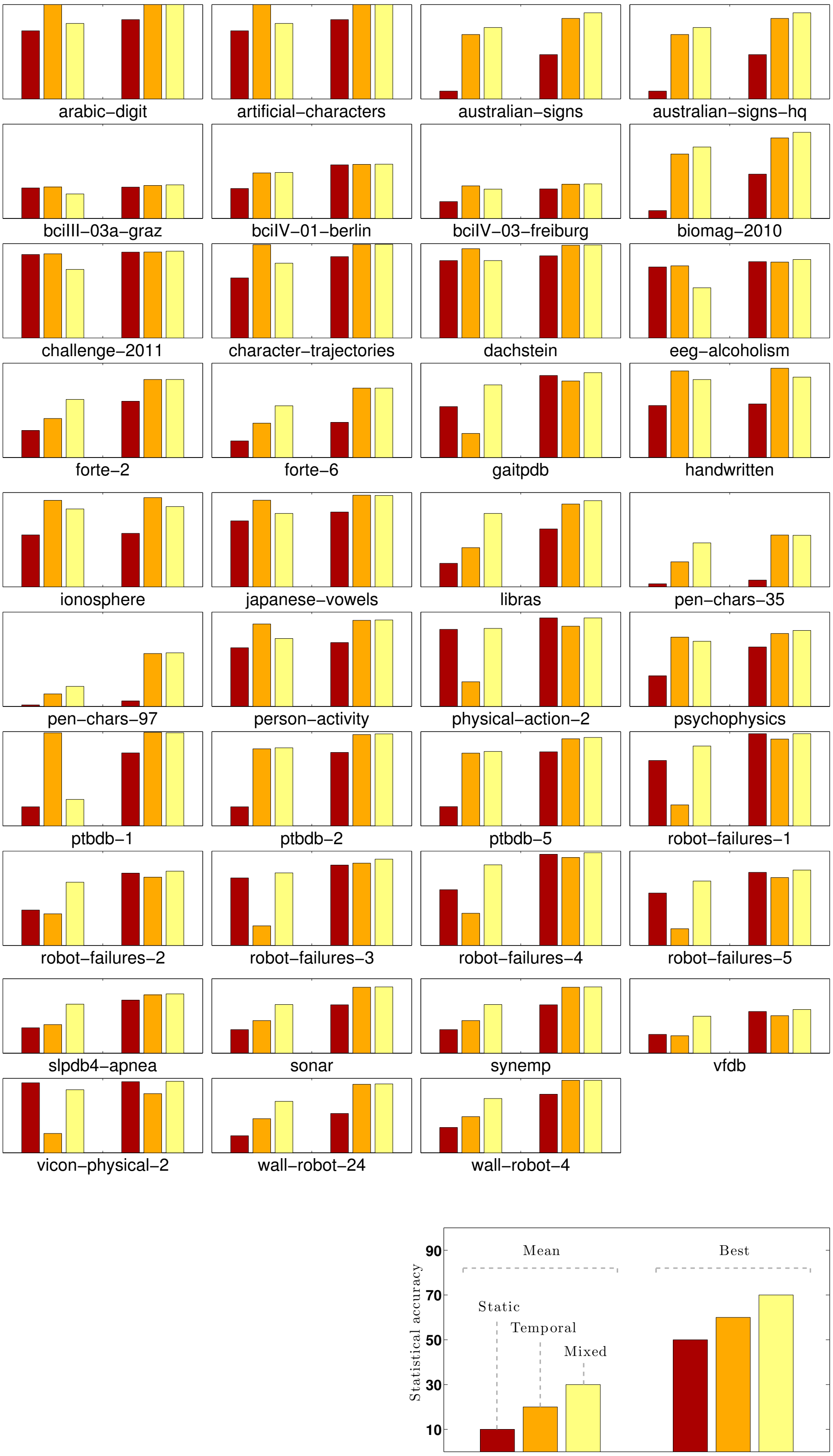



Figure 43: Comparison of the statistical significance for different parameters of warping for the DTW distance as opposed to simple resampling factors of the time series compared with the Euclidean distance.

### 9.5.3 *Warping or resampling*

We analyze in this section the influence of the warping window and resampling factors over the classification accuracy. Figure 44 presents the dataset-wise classification accuracy for the same time series given different warping window and resampling factors.

We provide for each dataset, the range of statistical significance for the *best* warping window and resampling factor (horizontal dotted lines). These results seem to indicate that, in most cases, the DTW can be outperformed by performing a coarse resampling of the series. This could be explained by the fact that performing such coarse quantification amounts to an implicit warping between series. However, it should be noted that we used an approximate computation of the DTW in order to reduce its quadratic complexity. Hence, the final accuracy results may have come superior by using a complete DTW computation. Nevertheless, if we focus on the differences of results for the resampling factors, an interesting distinction appears. It seems that three *types* of datasets emerge from these results. First, part of the dataset results seem almost unaffected by the modifications in resampling factors. Then, two opposite requirements seem to be imposed on the time series. Indeed, some of the datasets seems to require the most detailled information as possible to provide the best accuracy results. This can be seen in the evolution of accuracy for decreasing resampling factors in the *australian-signs, biomag-2010, japanese-vowels, pen-chars* and *ptbdb* datasets. This could be partly explained by the underlying complexity of the time series and also by the cardinality of classes. Indeed, if more classes are represented and the data is more complex, a more precise measurement of similarity is required. Oppositely, some of the datasets obtain their best results with very coarse resampling factors. This can be seen in the *bciIII-03a-graz, challenge-2011, handwritten, ionosphere, psychophysics* and *vicon-physical* datasets. In these cases, the use of more *prototypical* time series enhances the results. This could be explained by noise and outliers in the corresponding series. Indeed, performing a large downsampling implicitly provides a noise filtering mechanism.


Figure 44: Comparison of the statistical significance for different parameters of warping for the DTW distance as opposed to simple resampling factors of the time series compared with the Euclidean distance.

# 10 UNICITY OF HEART SOUNDS

As every organ, the human heart follows a physiological development singular to each living being. Its formation along the gestation steps of prenatal development makes it the first functional organ in an embryo. We show that, based on the HV-MOTS classifier, we can construct the first system that accurately identifies someone through the sounds its heart produces. We attain this goal by considering listening as an art and finding inspiration in musical analysis. This system is able to attain error rates equivalent to established biometric traits such as speech or gait recognition. We show *how* to listen to the peculiarities of each heart by developing a specific set of features based on the Stockwell transform, called the *S-Features*. Our findings are supported by the largest PhonoCardioGram (PCG) dataset ever collected. This includes the Mars500 isolation study of the Russian, Chinese and European spatial agencies. This dataset contains heart sound recordings that were recurrently collected over a time span of almost two years. The complete set of data allows to support biometric identification over large numbers of persons, long time spans and different physiological states. We also provide the first study ever on the phenomenon of template aging for heart signals.

## 10.1 BIOMETRIC SYSTEMS

Throughout its complete cycle, the heart produces a characteristic sound signature. Two major components usually emerge from these cardiac contractions. Listening to these sounds for the purpose of detecting anomalies in the circulatory system has been a common practice for the past two centuries. Hence, medical auscultation is taught as a form of art, which implies listening to the music of hearts with a firm clinical knowledge. Given this tradition, we could wonder if the sounds produced by a heart are utmostely *unique* to each individual. If we can exhibit this uniqueness through careful listening, we could be able to identify which person each heart belongs to with a single heart beat recording.

Therefore, we study the characteristics of heart sounds as biometric features. Biometric systems offer a natural and secure solution to authentication paradigms [180]. These methods seek to identify a person based on its distinguishing physiological or behavioral characteristics. The core strength of such systems is that the feature used for authentication is derived directly from the user. Therefore, it is unlikely to be lost, forged or stolen as it might be the case for *token-based authentication* (keys, passwords or cards).

A biometric system is essentially a *pattern matching* method. Hence, it establishes the authenticity of an identifying characteristic, by comparing it to a template collected in an enrollment procedure. Humans have always subconsciously applied such biometric recognition principles by analyzing the characteristics of the face, voice or even gait in order to identify other human beings. The most commonly known biometrics, like fingerprints or DNA, are now being used as wide spread international standards for identification with the emergence of biometric passports. Nevertheless, emerging biometrics are also being increasingly studied like facial thermogram, retinal scan, vocal prints or hand geometry. An ideal biometric feature should be *universal* throughout the population, *unique* to each person, *permanent* over time and easily *collectable* [180]. The biometric system itself should exhibit good *performance*, have a good *acceptability* by the population and prevent fraudulent attacks by *circumvention* [181].

Only very recently has emerged the idea of using the electrical activity of the heart (*ElectroCardioGram* (ECG)) as a biometric feature [46]. This information has been studied either through the direct extraction of representative points (called *fiducial points*) [343] or through spectral analysis [373]. Despite the nontriviality of using this information for identification, it has recently been shown that the features extracted from an ECG are invariant to the sensor location and physiological state of the subject [178]. The idea of using heart sounds should follow as a logical consequence [280]. However, studies in that line of research have been extremely limited and exhibited several flaws. First, most of the studies are evaluated on very narrow datasets, usually less than twenty subjects [117, 183, 280]. Some studies even make use of the same recording for enrollment and identification [281]. These methodologies cannot truely support any finding in a biometric context. Finally, the few studies that use larger databases with distinct recordings exhibit poor performances [37]. We provide in annex (cf. Section B.1) a complete comparison of existing biometric studies.

We show that we can access to the peculiarities of heart sounds if we consider listening as an art form and know *where*, *when*, *what* and *how* to listen. To know *where* we should listen, we study the unique frequency distribution of heart beats over a wide scope of subjects. To know *when* to listen, we use a segmentation procedure in order to focus our listening on each heartbeat separately. To know *what* to listen, we develop a specifically-tailored set of features based on the Stockwell Transform and inspired by *Music Information Retrieval* (MIR). Finally, to know *how* to listen we use the HV-MOTS classification paradigm.

## 10.2 heart physiology

The human heart is the muscular pump that manage the blood circulation throughout every parts of our body. The mechanisms and actions of the heart are distributed over its left and right halves. The right part dispatches the blood to the lungs and therefore dispenses pulmonary circulation. The left heart provides the supply of oxygen and nutrients to our entire body. Each half is further divided into two cavities known as the *atrium* and *ventricle*, where a set of valves (*tricuspid*, *mitral*, *pulmonary* and *aortic*) regulates our blood flow. Each of these components contribute in a specific way to the complete cardiac cycle along the periods of muscle contraction (*systole*) and relaxation (*diastole*).

Throughout its complete cycle (illustrated in Figure 45), the heart produces a characteristic sound signature constituted of two major components. There is no consensus on the precise physiological origins of every of their acoustic components. Some potential origins might be valvular (atrioventricular and exhaust valves), muscular (ventricular muscle) and circulatory (intracardiac blood flow) actions [238, 251, 354]. Even within this mixity, some major contributors have been clearly established through simultaneous PCG and echocardiography recordings. The first sound (*S1*) partly originates from the closure of the mitral valve and the subsequent (8 to 12 milliseconds later) opening of the aortic valve at the end of the isovolumic contraction. The second sound (*S2*) is induced by the closure of the aortic valve after the period of isovolumic relaxation. The onset of the S2 sound coincide the end of systolic contraction and the occurence of the

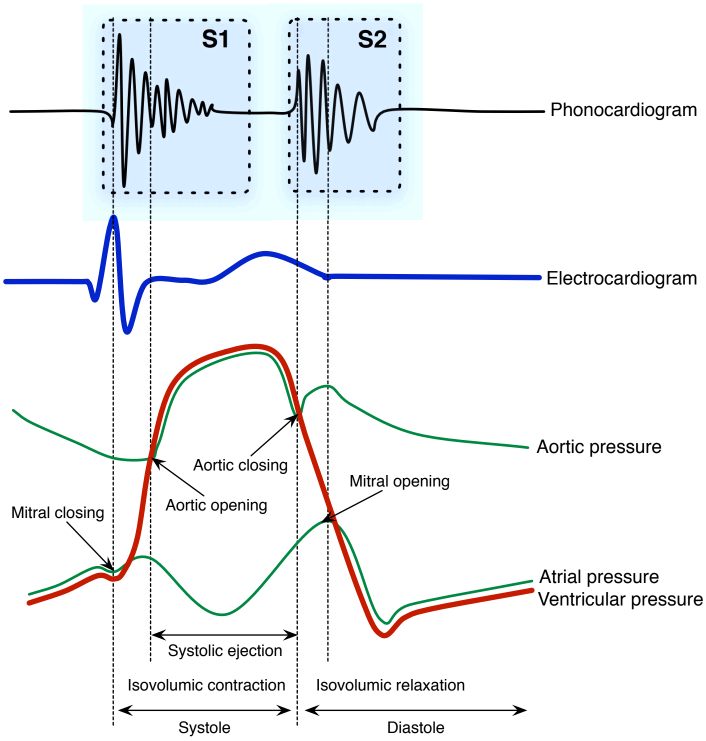



Figure 45: A complete cardiac cycle analyzed through recordings of the heart sounds (PCG), its skin electrical activity (ECG) and pressure in the aortic and atrial valves.

pulmonary ejection. The remaining period of diastolic rest is normally not audible in healthy subjects.

## 10.3 LISTENING TO THE HEART

Cardiac sounds are cyclic and therefore repetitive phenomena. A single cardiac cycle can thus be considered as the elementary unit of this study. We will therefore focus on listening from the beginning of systole, to the end of diastole. We will show that this period is consistent between several cycles of an individual by using the workflow displayed in Figure 46.

### 10.3.1 Where *to listen (pre-processing)*

As heart sounds possess highly specific characteristics, they require very cautious analysis. First, we need to uncover the useful bandwidth of cardiac sounds. Hence, we will be able to focus our listening on the range where the interesting spectral information lies. Figure 47 displays the typical frequency distribution of human heart sounds. This distribution is the normalized energy of the spectral information for 15.814 complete cardiac cycles recorded from 212 different subjects. We can see that the cardiac sound information primarily lies between 5 to 400 Hz, which makes it an significantly low-frequency signal. Therefore, we first process the heart signals with a [5,

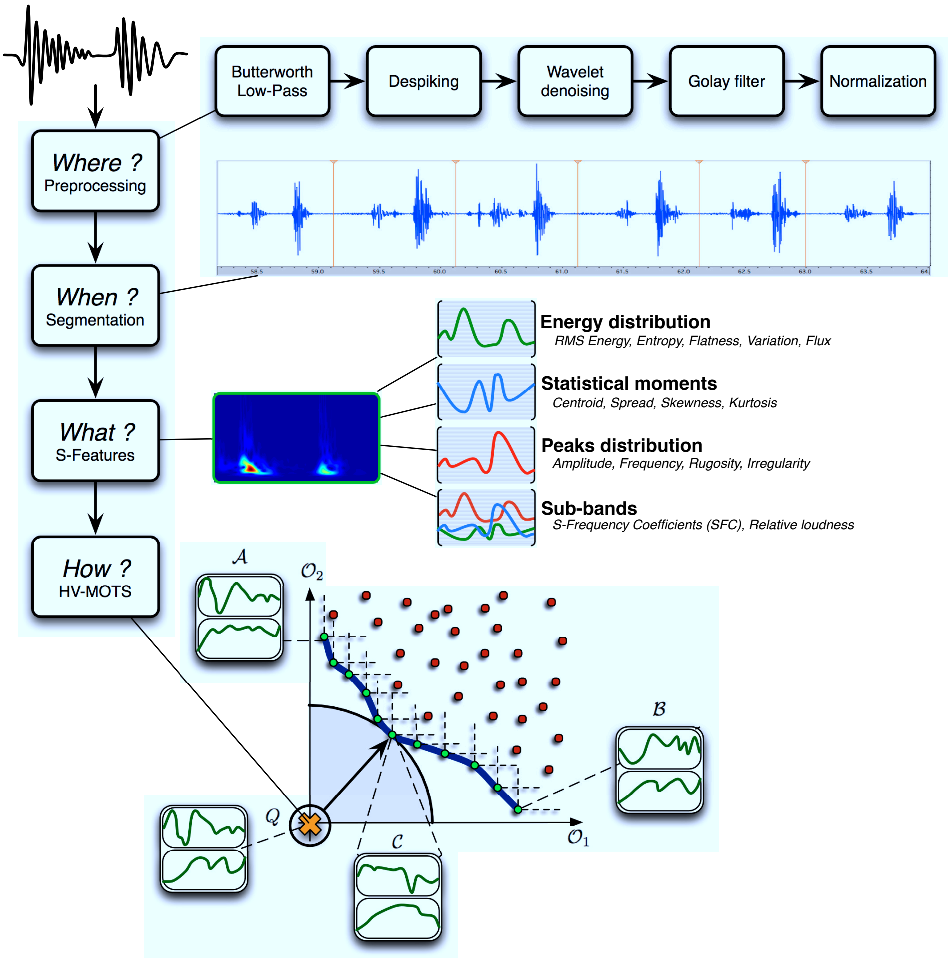



Figure 46: Algorithmic workflow for our heart sounds biometry system, summarizing the four milestone of listening which are *where*, *when*, *what* and *how* to listen.

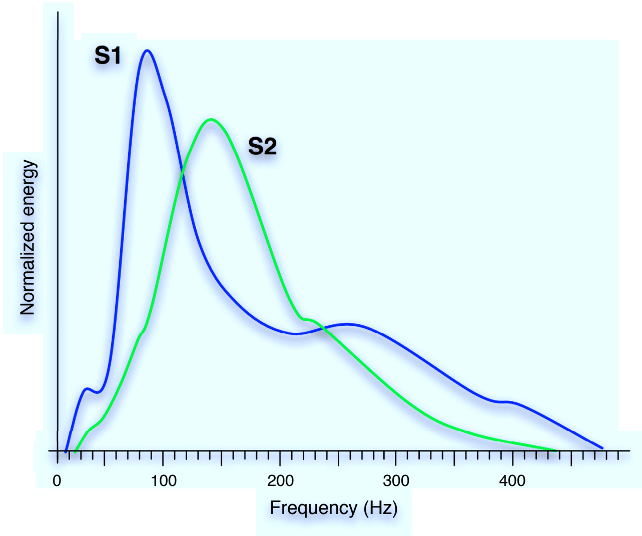



Figure 47: Typical frequency distribution of heart sounds based on the Stockwell transform analysis of 15.814 complete cardiac cycles. The S1 and S2 sounds recorded from 212 different subjects have been processed separately.

500]Hz band-pass filter, in order to remove all non-relevant information. As the present study is centered on the *sounds* produced by human hearts, it is subject to the same constraints as any audio-related system. Therefore, we need to remove the artifacts caused by noisy and low-quality recording conditions. These shortcomings are usually encountered due to the levels of surrounding noise but also the jitters caused by defects in the recording device itself. In order to alleviate these problems, we first process the signal with a despiking algorithm based on a phase space decomposition. Then, we apply a wavelet denoising algorithm with the 6th Daubechies wavelet (because of its similarity with the PCG signal). We can expect widely varying loudness levels in the resulting signals. Therefore, we apply a Golay filter of degree 9 with a window of 65 samples and an amplitude normalization procedure. This allows to enhance the signal of each heart beat. We provide in annex of this document (cf. Section B.3) a complete analysis of the impact from each pre-processing component on the overall performances.

### 10.3.2 When *to listen (segmentation)*

As we intend to use single cardiac cycles as elementary units, we need to precisely determine the boundaries of each heart beat. Therefore, we will focus on listening from the beginning of systole, to the end of diastole. Once these segments hae been detected, we can extract independent cycles for subsequent analysis. Therefore, we perform a transient segmentation based on the difference in spectral flux between successive frames. An energy variation threshold then allows to filter less significant transients. We apply an analysis window of 0.7 seconds with an 8 times oversampling. We extract and normalize each cardiac cycle and then store the resulting signals in separate files. Therefore, the identification template for each individual is provided by the complete set of cardiac cycles extracted from a single recording. We also provide in annex of

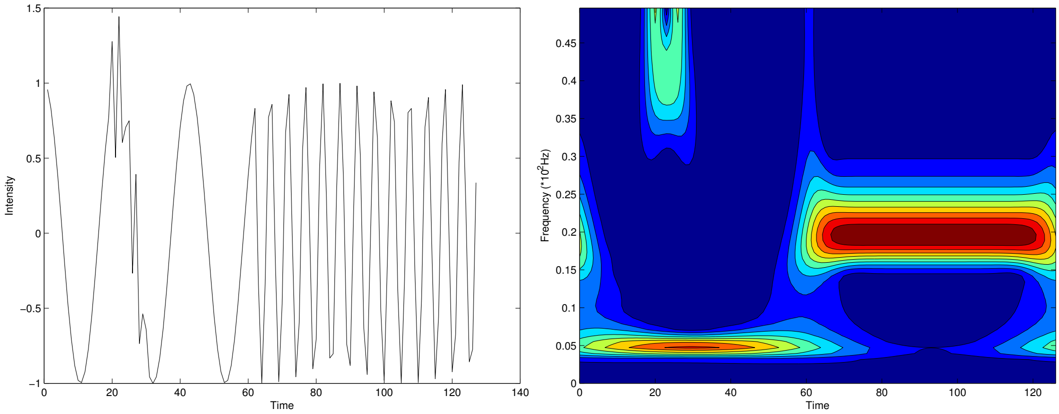



Figure 48: Illustration of the temporal and frequency resolution of the Stockwell transform. A synthetic signal (left) composed of two successive low-frequency sinusoidal components with a high frequency interference and their corresponding S-Transform spectrum (right).

this document (Section B.3) an extensive analysis of the influence of each segmentation parameter on the performance of the system.

### 10.3.3 What *to listen (S-Features)*

*The Stockwell transform*

The considerably low-frequency properties of heart sounds require a spectral transform that could adapt to these unique characteristics. Various signal processing tools have been developed to access the spectral information, such as the Short-Time Fourier Transform (STFT) or Continuous Wavelet Transform (CWT). Unfortunately, heart sounds have an extremely narrow bandwith with most of their energy radiating below the frequency resolution of these traditionnal decompositions. The Stockwell transform (or *S-Transform*), originally developed for analyzing geophysical data [338] provides an adequate solution to this problem. It is defined as a generalization of both the STFT and the CWT and overcomes some of their limitations. The S-transform of a function $h(t)$ can be defined as a CWT with a Gaussian mother wavelet multiplied by a phase factor [338]

$$S_x(t,f) = \int_{-\infty}^{\infty} h(\tau) \frac{|f|}{\sqrt{2\pi}} e^{-\frac{(t-\tau)^2 f^2}{2}} e^{-i2\pi f\tau} d\tau \tag{10.1}$$

The S-Transform exhibits a frequency-dependent resolution. This leads to an extremely fine resolution even at very low frequencies (below 50Hz) where lies the cardiac bandwidth. Therefore, it allows a better distinction of spectral components relevant to the cardiac information. Furthermore, unlike the CWT, modulation sinusoids are fixed with respect to the time axis. This property localizes dilations and translations and thus provides the same temporal resolution for every frequency bins. Figure 48 illustrates these properties. We use the fast S-Transform algorithm proposed in [60] which strongly reduces its computational complexity.

*S-Features*

The S-Transform provides an highly precise representation of the temporal evolution of frequency distributions for each cardiac cycle. However, the properties singular to each heart remain strongly diluted in such an overwhelming quantity of information. In order to uncover a potential uniqueness, we need to focus our listening by extracting higher-level information. To that end, we developed a specifically-tailored set of high-level features (called *S-Features*) based on recent research in music analysis (MIR). For the sake of clarity, we only provide here a summary of the implemented features but the complete mathematical definitions are available in annex of this document (cf. Section 10.3.3). First, we study the evolution of the S-Transform distributions by computing their statistical moments. Therefore, we calculate the mean (*centroid*), variance (*spread*), symmetry (*skewness*) and peakedness (*kurtosis*) of the successive frequency distributions. We further study the evolution of these distributions through their *energy*, *entropy*, *brightness, flatness, rolloff* and *variation*. We adapt the Mel-Frequency Cepstral Coefficients (MFCCs) for describing the evolution of spectral shape over various frequency bands through *S-Frequency Coefficients* (SFCs) (we detail this feature in the next section). Finally we study the evolution of peaks in the distribution by computing the relative differences in their frequencies (*roughness*) and amplitudes (*irregularity*). We resample each feature to a fixed length which allows us to abstract from the heart rate and therefore accounts for various physiological conditions.

*S-Frequency Coefficients (SFC)*

The S-Frequency Coefficients (SFCs) provide a description of the evolution of spectral shape over various frequency bands. Therefore, the computation of the SFCs follows a computing scheme analoguous to the MFCCs. However, the MFCCs process the spectral components through a filterbank based on the Mel scale in order to approximate perceptual results in timbre judgements. However, in the case of heart sounds, the Mel scale is not as relevant. Therefore, we use a logarithmic filterbank with specific distributions for the filters center frequency. If we look at the distribution of heart sounds (figure 47), we can see that most of the energy is concentrated under 200Hz. The overall bandwidth lies between 10Hz and 500Hz. Therefore, we compute the frequency centers

$$f^{i}_{center} = \left(f^{N_{bands}}/f^{1}\right)^{\frac{i-1}{N_{bands}-1}} .f^{1}$$

given a variable number of bands $N_{bands}$ and $f^1 = 10$ the center of the first filter and $f^N$ the center of the last filter. Note that this design will lead to extremely narrow filters in the low frequency range. Using such filters is possible with the S-Transform which provides one bin per frequency value. Finaly, a Discrete Cosine Transform (DCT) is used given its energy compaction property to reduce the dimensionality. We also compute the first derivative (DSFC) and second derivative (DDSFC) of the SFCs.

### 10.3.4 *How to listen (HV-MOTS Scoring)*

As we will discuss later (Section 10.4.2), the input evaluation required by biometrics system is quite different from classification tasks. Indeed, an identification attempt might be rejected if no template in the database match the input. Therefore, a biometric system should be able to make such a decision. To provide this subtlety, we maintain

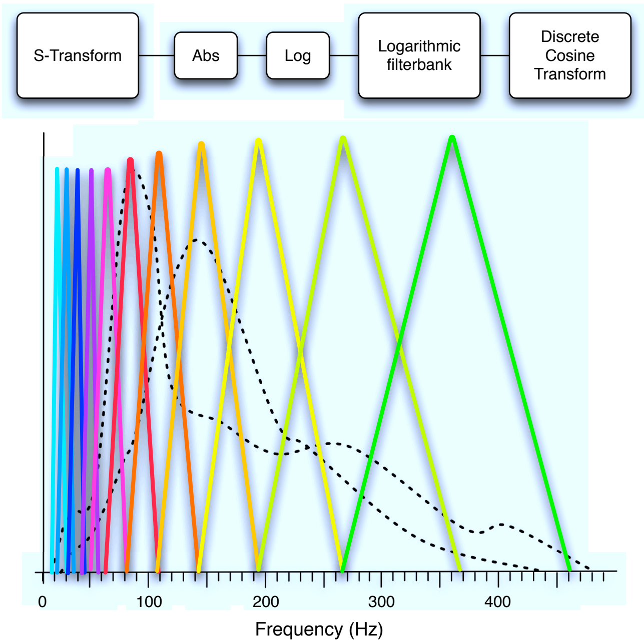



Figure 49: Computation workflow (up) and specific filter design (down) of the S-Frequency Coefficients (SFC). As we can see, the SFC are computed on a model similar to the MFCC. However, its fundamental differences comes from the use of the S-Transform and its filterbank designed to match the properties of heart sounds.

the workflow of HV-MOTS classification until the final decision. Hence, we do not select a particular class but rather keep the *normalized hypervolume* of each class as a matching score.

## 10.4 EXPERIMENTS

### 10.4.1 *Datasets*

We collected two datasets of PCG recordings, each of them having inherent advantages and flaws. Both datasets meet the main requirement to provide at least two distinct recordings for each person, collected at different times in separate sessions. First, the HSCT-11 dataset collected by Beritelli and Spadaccini [37] is the largest PCG set available of two distinct recordings per individual. This collection allows to assess the feasibility of heart sounds biometry over a large number of subjects. Nevertheless, the biggest flaw of this dataset is that it covers restricted time spans, most of the recordings being collected on the same day a few hours apart. Moreover, these recordings were collected from subjects in resting states only, which obviously can not account for varying physiological conditions. To overcome these shortcomings, we also include in our study a unique PCG dataset collected throughout the Mars 500 isolation experiment. This study is a joint Russian, European and Chinese spatial agencies psycho-social experiment. It was designed to analyze the psychological and physiological effects of a long-term deep space mission. Hence, it reproduces the conditions of a complete manned roundtrip to Mars. This dataset is the first of its kind for cardiac studies, and its advantages in our context are manyfold. First, its uniqueness stems from

| | Gender | | Ages | | | Records separation | | | Physical |
|---|---|---|---|---|---|---|---|---|---|
| | M | F | Min | Max | Med | Min | Max | Mean | state |
| **HSCT-11** | 157 | 49 | 15 | 96 | 28 | 1 | 2 | 1.1 | Rest |
| **Mars500** | 6 | 0 | 26 | 38 | 31 | 19 | 572 | 236.5 | Mixed |
| **Both** | 163 | 49 | 15 | 96 | 28 | 1 | 572 | 44.75 | Mixed |

Table 6: Details of the PCG recordings datasets. We provide the

the time spans over which auscultations have been performed. Indeed, PCG signals have been consistently recorded every three months over a complete period of almost two years of auscultation. This makes it the most detailled and longest systematical examination of individual heart sound recordings. Therefore, this dataset can unveil long-term evolutions and variabilities of heart sounds. Longer time intervals usually raise the difficulty in matching samples due to the phenomenon known as *template ageing*. This provokes an increase in error rates caused by time-related changes in the biometric pattern. We will therefore provide a specific evaluation framework designed to assess this effect. In addition, recordings have been performed under varying physical conditions which allows to account for fluctuating heart rates and diverse physiological factors. Moreover, recordings were collected by *auto-auscultation* from different non-experts in the medical field. This implies inherent *presentation* and *channel effects* with various levels of noise, placement of stethoscope and so forth. This also allows to account for the usability of such an identification system by novice users. All these characteristics provide an interesting range of robustness issues on the nature and variability of heart sounds. However, the Mars 500 dataset is flawed by its cardinality of six volunteers only. Therefore, we combine both datasets to evaluate the performance of our method. We follow the guidelines of [241] by providing a detailed description of the datasets. Table 6 summarizes with the demographics (gender and age) of the volunteers, time separation between samples and physical states under which recordings are made. Both dataset were collected with *cooperative*, *non-habituated* and *private* users with an *overt* and *attended* capture by an *open* system in a *standard* environment.

### *HSCT-11*

This database contains heart sounds acquired from 206 subjects (157 males and 49 females). The files were recorded in Wave format using a sampling frequency of 11025 Hz with 16 bits per sample. Two separate recordings were collected from each person (the length of the recordings varying from 20 to 70 seconds). Both recordings were usually collected the same day, separated by a break. Every person was sitting, in resting state while the auscultation was performed near the pulmonary valve.

### *Mars500*

This dataset is part of the *CardioPsy* study conducted in the Mars500 study. The hardware used was a digital stethoscope developed by the INFRAL society. The sounds are recorded by a piezoelectric microphone with linear response between 20 and 4000 Hz, AD converted and transmitted with Bluetooth to a laptop as hosting device. There is no pre-filtering of the sounds which are recorded at an 8000Hz sampling frequency with 16 bits per sample in Wave format. The dataset contains 54 sounds from the

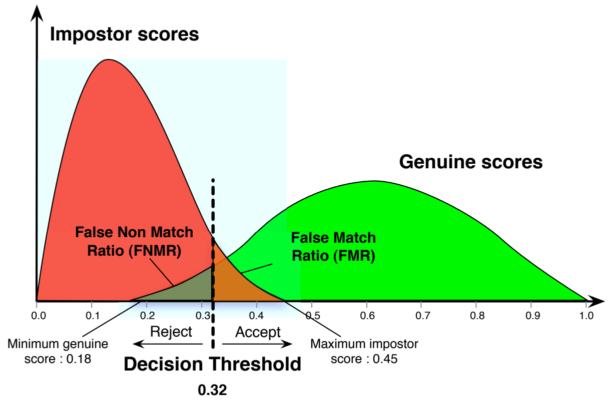



Figure 50: The evaluation of a biometric system in a real-life scenario given its distributions of *genuine* and *impostor* scores.

six volunteers, corresponding to nine cardiac auscultations that have been collected periodically over 520 days.

### 10.4.2 *Evaluation methodology*

Biometric identification systems strongly differ from classification problems. Indeed, in "pure classification" tasks, an input is bound to be labeled as belonging to one of the classes. Oppositely, a biometric system should also be able to detect if an incoming attempt does not belong to its template database. Therefore, the matching algorithm should decide whether the input is a *genuine* or *impostor* attempt given a set of reference individuals. To make this decision, the matching algorithms usually rely on a similarity score and a decision threshold (DT). If the score is inferior to the DT, the identification is negative (*impostor* attempt). Reciprocally, if the score is superior to the DT, the identification is positive (*genuine* attempt). Hence, biometric systems are prone to two types of errors. First, when the algorithm confuses an impostor and identifies it as a genuine attempt from someone inside the database. This type of error is called a *False Match* (FM). Second, when the algorithm fails to identify a genuine attempt (ie. fails to find a person which is truely enrolled in the database). This type of error is called a *False Non-Match* (FNM). There is a tradeoff between these two kinds of errors depending on the setting of the DT. Raising the DT reduces the number of FM errors but increases the number of FNM errors and vice versa. These concepts are illustrated in Figure 50.

Hence, in order to evaluate our proposal, the first recording is used for computing the reference templates stored in the database. Every further recording is used to compute the set of matching scores. For each of the $N$ persons in the database, the system computes one genuine score, and $N-1$ impostor scores. This yields a final number of $N$ genuine and $N.(N-1)$ impostor matching scores. As proposed by [145], we follow a thorough evaluation scheme and perform a multi-order analysis of results. This allows to obtain a comprehensive evaluation of the results from different viewpoints

| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
|---|---|---|---|---|---|---|---|---|---|
| $G_\mu$ | 0.811 | 0.830 | 0.862 | 0.889 | 0.912 | 0.930 | **0.933** | 0.931 | 0.917 |
| $G_\sigma$ | 0.014 | 0.017 | 0.018 | 0.018 | 0.018 | 0.019 | **0.019** | 0.021 | 0.017 |
| $I_\mu$ | 0.752 | 0.737 | 0.735 | 0.733 | 0.731 | 0.729 | **0.728** | 0.730 | 0.735 |
| $I_\sigma$ | 0.162 | 0.175 | 0.176 | 0.178 | 0.179 | 0.179 | **0.179** | 0.181 | 0.193 |

Table 7: Result of *Order-0 analysis* for an increasing number of objectives. This first columns display the distributions of genuine scores through their *mean* ($G_\mu$) and *variance* ($G_\sigma$). The following columns provide the same information for the impostor *mean* ($I_\mu$) and *variance* ($I_\mu$).

### 10.4.3 *Results*

We present here the results of identification over a combination of both datasets. This evaluation allows to assess the feasability of biometric identification over a wide range of persons. We perform a multi-order analysis of results as proposed on the assessment of biometric systems [145]. Therefore, we try to uncover the relation that exist inside these results. Then, we provide the first study on the effect of very long time spans (or *template ageing*) for heart sounds, thanks to the Mars500 dataset. Finally, we provide a comparison of our proposal to current state-of-art systems for serveral other biometric features.

*Order-0 analysis*

We collect the *order-0 statistics* by computing the complete distributions genuine and impostor scores. Then, we compute the first statistical moments of these distributions, ie. their *mean* and *variance*. Table 7 summarizes these statistics for an increasing number of objectives.

As we can see, the distributions of genuine and impostor scores appear to be very close to each other. This could endorse some limitations in the subsequent results. Indeed, biometric systems heavily rely on the principle that these distributions should be maximally separated. If we look at the results more closely, we can see that the genuine scores provide a remarkably high value with a quite narrow variance, which are actually the optimal conditions for a biometric system. However, the main problem lays in the distribution of impostor scores, which also induce a high similarity score. Moreover, the corresponding variance imply that the different distributions will most certainly overlap. These observations indicate that the first drawback of the system is to provide high scores for impostors. In both cases, this situation improves with an increasing number of dimensions, but starts to saturate after 7 objectives. However, the impostor variance still exhibits a steady increase in variance. This could be explained by the more inclusive Pareto fronts in higher dimensions. In order to analyze these distributions further, we count the total number of FM and FNM errors. This allows to compute the cumulative measurements at fixed rates of the DT. Hence, the *False Match Rate* (FMR) gives the percentage of FM errors and the *False Non-Match Rate* (FNMR) gives the percentage of FNM errors for a given DT. Figure 51 shows the tradeoffs between the FMR and FNMR for different levels.

As we can see, the previous observations seem to be confirmed in the decision tradeoffs. Indeed, we see that the gain in performance is, at first, quite large between the first and second level. However, it seems that these improvements reach an upper

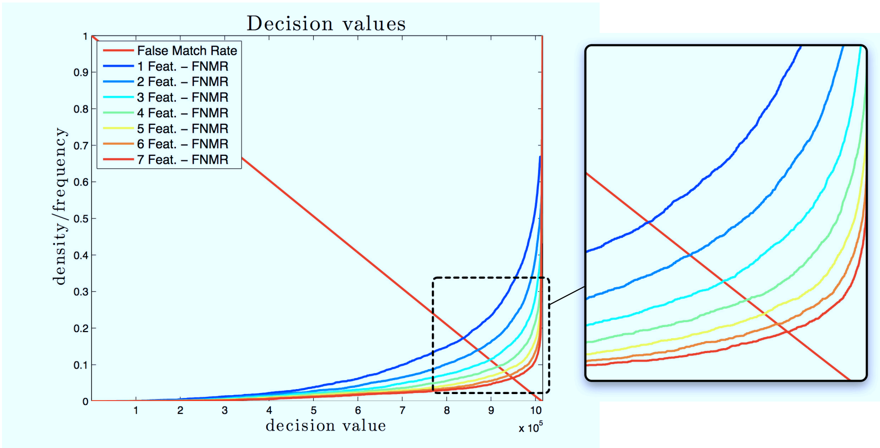



Figure 51: Decision tradeoffs between the False Match Rate (FMR) and the False Non Match Rate (FNMR) for an increasing number of objectives

| FNMR | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 0.1 | 0.251 | 0.173 | 0.123 | 0.093 | 0.073 | 0.059 | **0.051** | 0.056 | 0.058 |
| 0.01 | 0.566 | 0.422 | 0.306 | 0.229 | 0.175 | 0.141 | **0.116** | 0.119 | 0.134 |
| 0.001 | 0.669 | 0.540 | 0.416 | 0.318 | 0.255 | 0.206 | **0.171** | 0.178 | 0.188 |

Table 8: Evolution of *FMR* given a set of fixed *FNMR* for an increasing number of objectives.

bound around 6 dimensions. We did not include the results over 7 dimensions for the overall clarity of the figure. Furthermore, the distribution of decisions values seems to exhibit a strong skewness when higher number of objectives are involved. This result can be correlated with the distribution of impostor scores, which induces high values. Hence, the tradeoff between FMR and FNMR seems to reach an equilibrium only towards higher scores in both genuine and impostor.

*Order-1 analysis*

We analyze the *order-1 statistics* by computing the *trade-off values* between the FMR and FNMR over the complete range of decision thresholds. This allows to assess the performances of the system under different sets of constraints. Indeed, low FNMR are preferred for fast identification (usually forensics). Reciprocally, low FMR are mandatory for high security applications. Table 8 summarizes the evolution of *FMR* given a set of fixed *FNMR* for an increasing number of objectives.

As we can see, the results improve with a higher number of dimensions until they hit an upper-bound and regress. This correlates well with the observations made at the previous level. A first glance at the results indicates a low FMR for loosened FNMR constraints (0.1). However, the best error rates for strict FNMR constraints (0.001) are still too high for a high-security framework. Furthermore, another indication provided by this analysis is that the system seems more sensitive to FM than FNM errors. This can be observed by the evolution of the FMR errors at low FNMR (0.001) which seem to improve remarkably faster compared to higher FNMR constraints. To confirm this hypothesis, we provide the *Detection Error Trade-off* (DET) curve, which represents the co-evolution of FMR and FNMR by varying the DT. Figure 52 shows the *Detection Error*

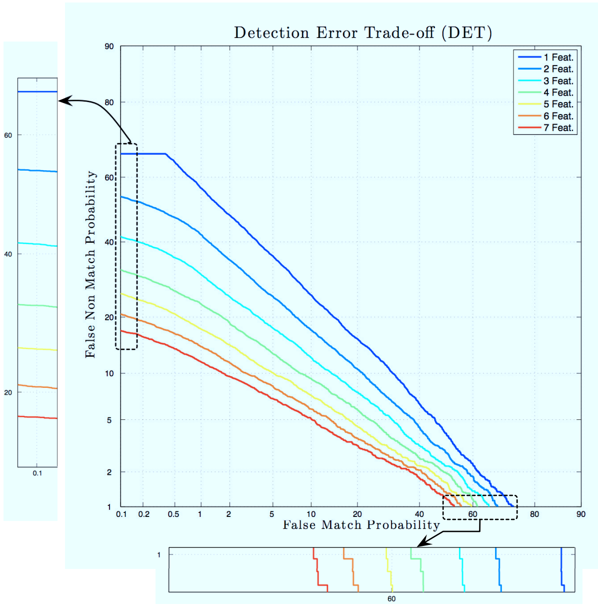



Figure 52: Detection Error Trade-off (DET) curves for increasing number of objectives.

*Trade-off* (DET) curves for an increasing number of objectives. The curves are plotted parametrically on a log-log scale.

As we can see, the overall distribution of the DET curves seems to confirm our previous hypotheses. An increasing number of objectives still improves the results until it reaches the previously exhibited upper-bound. However, there is a clear disparity on this improvement between FMR and FNMR. It seems, from the distributions of DET curves, that the system does not improve on FM errors as well as it improves on FNM errors. This would tend to confirm that the system is more sensitive to FM than FNM errors. Hence, these results could prevent the use of heart sounds biometry in high-security frameworks. However, this indicates that the system could be notably more interesting for high identification frameworks (such as forensics), as they require the reciprocal constraints.

### *Order-2 analysis*

The *Order-2 analysis* computes the previous statistics for all possible threshold values and the *Rank-1 identification rate* which is the number of times the correct person has the highest score. We also compute the *Equal Error Rate* (EER) which is the percentage of errors induced by a particular DT setting thanks to which the number of FM and FNM errors are equal. The EER provides a good summary of an algorithm performance on both types of errors. Finally, we compute the *Receiver Operator Characteristic* (ROC) curve, which is similar to DET curve, but plots the *True Acceptance Rate* (TAR = 1 - FNMR) against FMR. Table 9 summarizes the *EER* and *Area Under the ROC Curve (AUC)* for an increasing number of objectives.

As we can see, the results in EER confirm the observations made at previous orders. An increasing number of dimensions allows to provide a steady improvement for the 6 first dimensions. Then, the system exhibits an upper-bound in performances with an EER of 6.524% and starts to regress for higher number of dimensions. The AUC follows the same overall evolution. This final EER value seems to indicate that the biometric

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| EER | 17.27 | 13.98 | 11.34 | 9.499 | 8.088 | 7.095 | **6.524** | 6.735 | 6.788 |
| AUC | 0.909 | 0.936 | 0.953 | 0.964 | 0.972 | 0.976 | **0.980** | 0.978 | 0.972 |

Table 9: Result of the *EER* and *Area Under the ROC Curve (AUC)* for an increasing number of objectives using the HV-MOTS method

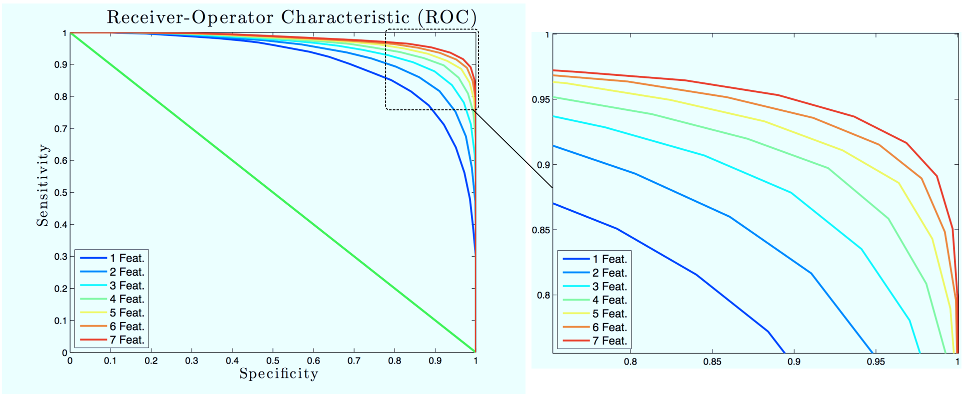



Figure 53: Receiver-Operator Characteristic (ROC) curve for an increasing number of dimensions

identification through heart sounds is feasible. Indeed, the overall error rate in both types of errors implies that the system will provide an erroneous answer only 6.524% of the time. This means that the system is amply over the random chance probability. We can see this evolution more clearly in Figure 53 with the *Receiver-Operator Characteristic* (ROC) curves for increasing number of dimensions.

### *Order-3 analysis*

Finally, we perform the *order-3 analysis* by plotting the *Cumulative Match Characteristic (CMC)* curve which displays the *Rank-k identification rates* for all possible ranks. We also compute the differences between the first and second best scores (only in the cases where the genuine attempt is scored as first). Table 10 summarizes the *Rank-1*, *Rank-2* and *Rank-5* identification rates, as well as the *mean* and *variance* of the first-to-second scores differences.

This analysis provides results that can be related to "*pure classification*" tasks. Indeed, the *Rank-1* identification rates can be seen as an analogous to classification accuracy as it provides the percentage of time that the true genuine score is in first position. It

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Rank-1 | 33.12 | 48.79 | 60.83 | 70.28 | 77.01 | 81.29 | 83.59 | 84.61 | **85.01** |
| Rank-2 | 41.98 | 57.63 | 68.94 | 77.03 | 82.34 | 85.85 | 87.20 | 88.12 | **88.76** |
| Rank-5 | 55.14 | 68.68 | 77.86 | 83.86 | 87.77 | 90.35 | 91.56 | 92.43 | **93.12** |
| $1^{st}2^{nd}$ $(\mu)$ | 0.009 | 0.012 | 0.014 | 0.017 | 0.019 | 0.021 | **0.023** | 0.021 | 0.021 |
| $1^{st}2^{nd}$ $(\sigma)$ | 0.043 | 0.071 | 0.073 | 0.076 | 0.078 | 0.081 | **0.083** | 0.082 | 0.079 |

Table 10: Result of *Order-3 analysis* for different levels using the HV-MOTS method

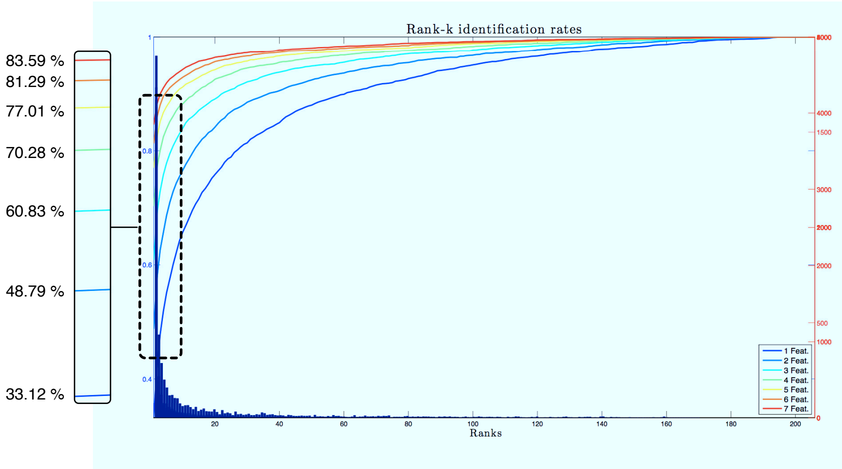



Figure 54: Results of the system identification based on the Rank-k identification rates

appears that these results do not follow the same evolution as previous orders. Indeed, we can see that growth in identification rates is continuous, even after 7 dimensions. This also exhibits the usefulness of the multi-order analysis of results, as it allows to detect subtleties in different viewpoints of the results. However, these rates still seem to progress less importantly after 6 dimensions are involved, which still confirms an upper-bound in results. Moreover, the previous observations can still be outlined in the differences between the first and second scores. Figure 54 provides the *Rank-k identification rates* for all possible ranks in order to enable a graphical comparison of the evolution of these results.

*Menagerie analysis*

The distribution of matching scores (either genuine or impostor) can exhibit *outliers* that need further investigation. As discussed by previous research [389], score distributions may exhibit the presence of user groups that are fundamentally different from the general population. Therefore, we study user variations by performing a *menagerie analysis* to assess the existence of users termed as *chameleons* (simultaneous high genuine and impostor scores), *phantoms* (simultaneous low genuine and impostor scores), *doves* (high genuine and low impostor scores) and *worms* (low genuine and high impostor scores). By exhibiting these singularities in users, we can further study these animals, which could exhibit weaknesses in the system. Figure 55 shows the results of the menagerie analysis for every heart beats in the dataset.

As we can see, the major concern for the system is the *phantoms* animals. These indicates that some of the heart beats provide very low genuine scores and also low impostor scores which makes them hard to identify correctly. Furthermore, some *worms* also appear in which the mean impostor score is significantly higher than their genuine scores. Both of these species will cause large errors in the evaluation of results. Moreover, it seems that these instances all belong to different, which will further distort the results. Finally, *chameleons* also appear in consequent quantities and are "double-sided" animals. Indeed, these animals have simultaneously high impostor and genuine scores. This implies that even though they may be correctly identify, they can alter the similarity scores of other heart beats. This can lead to increased error rates in the final evaluation. If we look at the overall distribution of these animals, it seems that

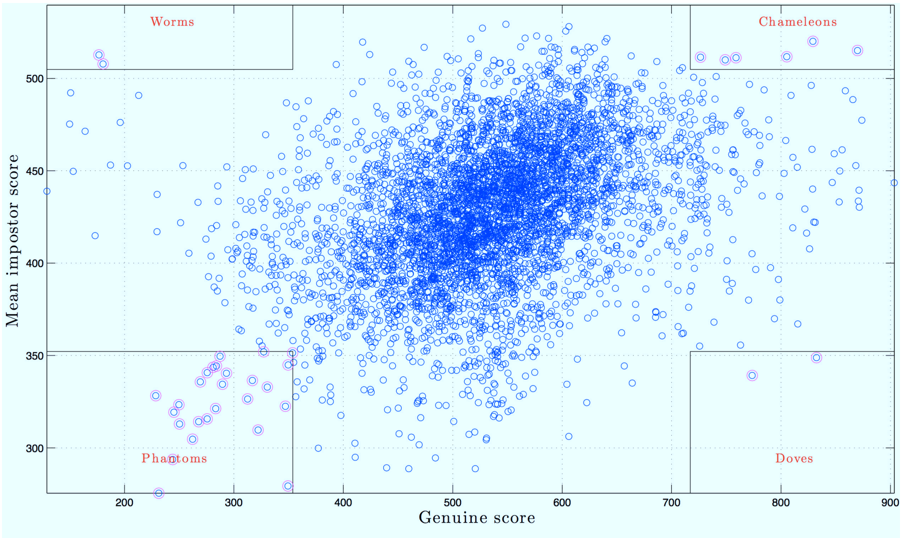



Figure 55: Results of the menagerie analysis performed over every heart beats for the best feature combination.

they appear sporadically for a wide range of persons. Hence, this could indicate some defects in the segmentation procedure.

*Comparison of datasets*

We compare the results of our proposal with previous results and between the two different datasets. Each of these collections has specificities that allow to study different aspects of robustness on our system. First, the HSCT-11 dataset allows to study the heart sounds biometric identification over large numbers of users. The baseline EER value for the HSCT-11 database is 13.66%, obtained using the UBM/GMM method described in [37]. Our method allows to obtain an EER of 6.524% which strongly outperforms previous research on this topic. This result confirms the feasibility of heart sounds identification. The Mars500 dataset even if limited in the number of persons has numerous interesting properties. Thanks to it, we now provide the first study ever performed on the *template ageing* for heart sounds biometry.

*Template ageing*

This phenomenon is known to be one of the major concern in biometrics and arguably one of the main potential weaknesses of the heart biometric feature. Indeed, the occurence of cardiac diseases is known to strongly impact the sounds produced by hearts. Furthermore, even healthy human hearts are bound to evolve with age but these long-term effects are still unknown. We provide a first analysis of this phenomenon thanks to the Mars500 dataset. For each of the six subjects, recordings have been performed reccurently every three months, leading to a collection of eight samples per subjects with a maximal separation of 572 days between recordings. Our methodology is therefore to compare every possible pairs of recordings and to create *classes of ageing* for similar separations between different recordings. This leads to a total of seven ageing classes, each of them being three months further apart. Table 11 summarizes the

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| **Months** | 3 | 6 | 9 | 12 | 15 | 18 | 21 |
| **Instances** | 42 | 36 | 30 | 24 | 18 | 12 | 6 |
| **EER** | 5.50% | 1.00% | 12.43% | 2.50% | 13.66% | 12.76% | 5.33% |

Table 11: Results of the analysis of the *template ageing* phenomenon.

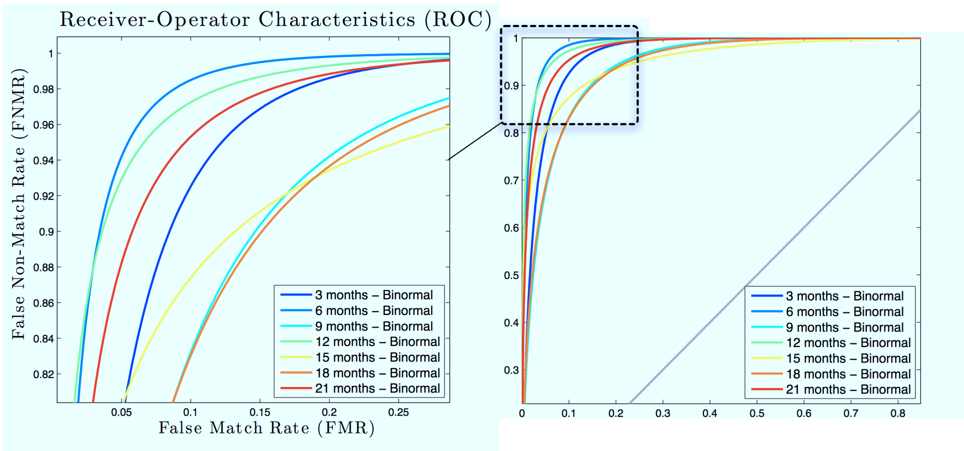



Figure 56: ROC curves for the analysis of the *template ageing* phenomenon. Because of the relative scarcity of instances, the curves have been fitted with a binormal function.

distribution of these samples, number of testing instances and overall EER obtained for each class.

As we can see, the evolution of the EER is not correlated with the separation templates. Hence, the variations of EER with more distant templates can not truly be ascribed to the time spans between which they have been collected. Indeed, it seems that the best identification is provided inside the *6 months* class, followed by the *12 months* and *21 months* classes. It also appears that the same overall identification performance is provided by the 3 and 21 months of records separation. Therefore, the variations of EER could be more related to the sample themselves (quality of recordings) than their time frames. Indeed, we must recall that these were collected from *auto-auscultation* by *non-medical experts.* Therefore, this setting of (non-uniform) data collection can strongly influence the results of the system. However, these results seem to confirm a form of stability in the heart sounds signals over a time span of two years. These results also outline the extreme caution with which should be undertaken an experimental analysis of biometric systems. Indeed, we can see that by just selecting different parts of the same dataset, we can provide an EER that varies from 1.00 to 13.66%, which is a enormous gap in the performance of a system. Therefore, analysis of results should always be the more detailed and comprehensive as possible. Figure 56 provides the ROC curves for the graphical comparison of the *template ageing* phenomenon. The ROC of each *ageing class* is plotted independently. Because of the relative sparsity of available instances, the curves have been fitted with a binormal function.

*Parameters influence*

We evaluated the influence of all the parameters implied in our heart sounds biometry system. However, for the sake of clarity, the complete analysis of all the parameters is available in Annex (cf. Section B.3).

### 10.4.4 *Comparison to existing biometrics*

We present in this section the comparison between our proposal and current results for various biometric features. We tried to gather the most recent results of major competitions in biometrics. The results presented here are extracted from the *Fingerprint Verification Competition* (FVC) [109], the *Iris Exchange* (IREX) [190], the *Noisy Iris Challenge Evaluation* (NICE) [289] and the *National Institute of Standards and Technology* (NIST) [360]. We chose to compare our proposal to various biometrics based on the best EER value as it allows to perform a straightforward comparison of systems performances. It should be noted that for face recognition, the NIST recommends not to use the EER as evaluation measure as it makes the system work at unrealistic False Match Ratio.

| Biometrics | EER (%) | Genuine | Impostors | Reference |
|---|---|---|---|---|
| Iris | 0.057 | 6000 | 6000 | [190] |
| Fingerprint (Standard) | 0.108 | 27720 | 87990 | [109] |
| Fingerprint (Hard) | 0.687 | 19320 | 20850 | [109] |
| Iris (Noisy) | 1.31 | 1876 | 1876 | [289] |
| Palmprint | 2.569 | 2800 | 4950 | [70] |
| Face recognition | 2.83 | 943 | 88306 | [54] |
| Signature (Online) | 2.85 | 6374 | 9378 | [47] |
| Gait analysis | 6.2 | 50 | 2450 | [254] |
| **Heart sounds** | **6.524** | **212** | **44732** | **-** |
| Speech | 7.9 | 310 | 95790 | [360] |
| Signature (Offline) | 9.15 | 6527 | 9360 | [47] |
| Keystroke | 10.37 | 133 | 6732 | [138] |
| ECG | 10.8 | 73 | 5256 | [330] |
| Eye movement | 27.0 | 32 | 992 | [171] |

Hence, we can see that the heart sounds biometry performs better than several recently emerging biometric modalities. It seems that the heart system even performs better than speech biometry. Furthermore, even if it seems that gait analysis performs better than heart sounds, a special attention should be given on the number of genuine and impostors comparisons that are provided by the systems. Indeed, with only 50 genuine comparison, the gait analysis provides a somehow weaker panel of analysis. As we have seen in the previous section, this difference in cardinality can have a tremendeous impact on the final results. This weakness is less pronounced for heart sounds but can still be outlined as we will detail in the next section.

| Factor | Description of the effect |
|---|---|
| *Demographics* | Children can have rapid overall physiology changes whereas older people tends to have heart conditions more often. |
| *Template ageing* | Changes in the users biometric pattern can vary with wider delays between the enrolment and identification. |
| *User behavior* | Affects the recording process through unrequired movements |
| *Physiological* | The physiological state (*physical activity, stress, tension, relaxation*) might impact the heart sounds |
| *Environmental* | Disturbances in the recording process with background sounds (such as the subject speaking) or noises |
| *Sensor and hardware* | Differences in the sensor quality, the pressure applied and the transmission channel used to acquire the data. |
| *Heart diseases* | Changes produced by diseases can produce a wide disparity over time with both chronic and temporary illnesses. |
| *Spoof attacks* | Can hardly be performed on the system but should be considered like pre-recorded sound inputs |

Table 12: Lists of both user and environmental factors that could potentially affect the performances of the heart sound identification system.

## 10.5 DISCUSSION

We provide here a discussion on the advantages and drawbacks of heart sounds as biometric features. One of the inherent problem of studying heart recordings comes from their time varying nature. Indeed, commonly used biometrics like fingerprints or face recognition offers an utmost advantage in which the features are "time-static". This implies that the template can be extracted from a snapshot frozen in time. Oppositely, a heartbeat signal is inherently temporal and therefore needs time to develop. Furthermore, the system requires at least a few heartbeats to obtain better performances which therefore requires more acquisition time. When comparing the enrolment template to the identification input, their differences may also be amplified by several external factors. We provide in Table 12 a list of these factors. Therefore, we could argue on our methodology that the disparity of data is too thin. Indeed, the datasets do not fully cover the problems of physiological conditions (rest, stress, anxiety, exhaustion) that may impact the morphology of the heart sounds. The topic of heart diseases has been willingly left aside, even if it could be one of the main shortcomings of our system. Indeed, a medical condition can strongly disrupt the sound signature of the same person's heart. Furthermore, even if we provided a first glance on the problem of template ageing, we still do not know the effect of this phenomenon on the scale of several decades. Another problem of interest emerge from the recording process itself. Indeed, several auscultation beaches can be used to obtain a recording. The heart sound signature being filtered by the surrounding body, the choice of these beaches may also influence the identification results. Finally, because of its emerging nature, the study of heart sounds distinctiveness suffers from an evident lack of data. Testing the scalability of our study would require a larger population of study.

So we could wonder why the biometric identification is working even within all these limitations. We can provide a first answer to this question by taking a closer look at the mandatory properties of biometric features [180]. First, regarding the *uniqueness* property, it is supported by medical evidence that the physiological variability (like mass distribution) provide each heart signal with distinctive characteristics [361]. Moreover, medical research has even been devoted to reduce the variability in heart signals between individuals for diagnosis purposes [112]. In this regards, we can hypothesize that the development of the flesh of hearts is unique to each person but also the development of surroundings areas (bones, lungs, skin). This also provides a first sketch on the success of the S-FC feature as the heart sounds can be related to a source-filter paradigm, equivalent to speech (in which case the MFCC are known to be a weapon of choice). Furthermore, using the heart sounds as biometrics shows no flaw in the *universality* property as a beating heart is quite mandatory for any living person. Moreover, recent results [4] suggests that medical biometrics is not only *permanent* over a long period, but also allows continuous identification with successive recordings input to the system. We exhibited this permanence in a biometric context which also show that the heart signal is consistent over a span of at least two years. Heart sounds also offers a very low *circumvention* and should be extremely robust to spoof attacks. Indeed, it seems hardly imaginable to forge or stole an heart as it undoubtedly need its possessor to produce a sound. Furthermore, it seems also hard to be concealed or hidden as burning fingerprints. Finally, the remaining properties of *collectability* and *acceptability* would require large scale user studies to provide a definitive answer.

# 11 AUDIO APPLICATIONS

The HV-MOTS classification framework was inspired by observations from our auditory perception. Therefore, it seems natural to now turn our attention to audio applications of our method.

## 11.1 GENERIC AUDIO CLASSIFICATION

### 11.1.1 *Content-based audio retrieval*

Content-based audio retrieval has become a popular research field, notably through the appearance of QBH introduced by Ghias et al. [137]. Most of researches devoted to this topic are based on symbolic song databases and therefore use the notion of *pitch contour* [358], which is the sequence of relative differences in pitch between successive notes. For content-based audio clips classification, the first system was proposed by Wold et. al [380]. In their study, the sounds were represented by a vector of mean, variance and autocorrelation values of spectral features. These vectors were then compared with the Euclidean distance as similarity metric. This approach known as *Bag-Of-Features* has been extended using larger sets of features [372] or adding *relevance feedback* [291] in which the user selects its preferred results for refinement [371]. Subramanya et al. [342] used frequency coefficients from spectral decompositions and showed the superiority of DCT. They later used the multi-resolution property of the wavelet transform [341] and showed its robustness to noise. However, the selection of coefficients yields very large vectors which may be unsustainable for massive datasets. This approach was extended in Li and Khokhar [222] by using multiple statistical values over wavelet coefficients. This allows hierarchical indexing, as proposed in Li and Hou [224] with a pyramidal algorithm which provides an acceleration over previous approaches. Several indexing and learning schemes have also been investigated. Li [223] proposed the Nearest Feature Line (NFL) based on the idea that in feature space, lines between similar audio clips represent continuous deformations between class properties. Therefore, comparisons with queries are made with these feature lines. However, computing NFLs between every sound samples seems to induce a large computing and storage overhead. Other machine learning techniques like Boosting [154] or GMM [165] were studied but they seem to be outperformed by the SVM-based approach.

Regarding temporal modelization, Cai et. al [64] proposed to use templates of temporal patterns for energy, harmonicity and pitch contour. Although they showed to improve accuracy, this approach seems hardly scalable because of the relative simplicity of the patterns used. More generic temporal modelization with HMMs [403] has been proposed, where comparison of HMM likelihoods with the query allows to obtain a ranked list of results. Casey [74] proposed to use the MPEG-7 feature set with an Independant Subspace Analysis (ISA) to obtain the most salient features of a sound. He further introduced a minimum entropy method [75] to train the HMM classifier which appears to outperform classical training. However, the ISA usually yields large computational overheads. The superiority of HMM cross-likelihood ratio has been

| Musical instruments | | Effects | |
|---|---|---|---|
| Altotrombone | 13 | Animals | 9 |
| Bells | 7 | Crowds | 4 |
| Cellobowed | 47 | Laughter | 7 |
| Oboe | 32 | Machines | 11 |
| Tublarbells | 19 | Percussion | 99 |
| Violinbowed | 45 | Telephone | 17 |
| Violinpizz | 40 | Water | 7 |
| **Speech** | | | |
| Female | 35 | Male | 17 |
| | | **Total** | 409 |

Table 13: Description of the MuscleFish dataset used in classification tasks. 409 sounds are divided into 16 classes.

shown over GMM [363] and feature histograms [163]. However these studies exhibited that all the approaches are very sensitive to noise and low-quality sounds.

### 11.1.2 *Datasets*

We evaluate the HV-MOTS paradigm for content-based audio classification using two datasets. These are organized following the same structure as presented earlier (Section 7.6). First, the reference MuscleFish dataset [380] allows to compare our approach to state-of-art methods. Second, we collected a more recent and comprehensive dataset to test how our approach scales up to wider sets of data. Both datasets are available on a supporting web page [1] so that the results of our experiments are fully reproducible.

*MuscleFish*

This dataset, assembled by Wold et. al [380], has been used extensively [154, 153, 223, 305, 325] in order to compare performances of different systems. It is composed of 409 sound files which are divided into 16 classes. Complete description of the dataset is presented in Table 13. Files are single-channel Sun/Next (.au) µ-law encoded audio files quantized to 8-bit with a sampling rate of 8 kHz. Loudness levels and file lengths vary over samples with the average size of a file being about 50 KBytes.

*Freesound*

In order to evaluate how our approach scales up to more comprehensive datasets, we collected 2193 sounds representing 54 classes from the Freesound project [2], which makes this set five times larger than the MuscleFish dataset. Complete description is presented in Table 14. Files are single and double channels, WAVE and AIFF format, quantized to a minimum resolution of 16-bit with a minimum sampling rate of 44.1 kHz. Loudness levels and file lengths vary with the average size of a file being about

1 http://repmus.ircam.fr/esling/ieee-mots.html
2 http://www.freesound.org

| **Western instruments** | | **Indian instruments** | | **Animals** | | **Effects** | |
|---|---|---|---|---|---|---|---|
| Alto-flute | 85 | Santoor | 15 | Birds | 29 | Applause | 41 |
| Bassoon | 80 | Singing-Bowl | 8 | Cat | 26 | Footsteps | 33 |
| Cello | 59 | Tabla | 18 | Dog | 21 | Gunshots | 15 |
| Clarinet | 76 | Tambura | 16 | Horse | 18 | Heartbeats | 23 |
| Contrabass | 66 | Thumb-piano | 17 | **Synthesis** | | Laughter | 141 |
| Glockenspiel | 17 | **Drums** | | Bassline | 99 | Musicbox | 28 |
| Guitar (dist) | 16 | Crash | 43 | Reese | 47 | Paper | 17 |
| Oboe | 113 | Hi-hats | 25 | Vocoder | 43 | Siren | 37 |
| Saxophone | 27 | Kick | 27 | Wobble | 39 | Subway | 11 |
| Trombone | 42 | Loops | 25 | **Speech** | | Sword | 21 |
| Trumpet | 35 | Snare | 42 | Female | 96 | Telephone | 16 |
| Tuba | 74 | Scratch | 20 | Male | 87 | Thunder | 29 |
| Viola | 58 | Toms | 9 | Robotic | 31 | Water | 43 |
| Violin | 98 | Tone | 8 | Scream | 32 | Whistle | 26 |
| | | | | | | Zipper | 17 |
| | | | | | | **Total** | 2193 |

Table 14: Description of the Freesound dataset collected specifically for our study. 2193 sounds are divided into 54 classes.

310 KBytes. One particularity of this dataset is that it includes a section for synthesis sounds, which are classified based on their temporal morphology.

*Evaluation methodology*

We use the same *Leave-One-Out* evaluation methodology as in the previous chapter. We measure the *classification accuracy* by testing combinatorial possibilities among available descriptors. Hence, we evalute the classification for every single descriptor listed in Table 2 then every combination of two descriptors, and so forth. As previously, we keep only the top performing half of the descriptors after each step, based on their classification accuracies.

### 11.1.3 *Results analysis*

We present in Figure 57 the classification accuracies on the MuscleFish dataset for a growing number of objectives. For a given number of objectives, the top figure provides the *mean* accuracy over every combination and the figure below is the *best* score obtained by a single combination. As we can see HV-MOTS consistently outperforms the other approaches in classification accuracy. This result is confirmed by the accuracies obtained on the Freesound dataset, presented in Figure 58. Even with up to five times more classes and sounds, HV-MOTS exhibits an almost equivalent classification accuracy and still outperforms other methods.

More interestingly, it seems that HV-MOTS strongly outperforms other approaches in *mean* classification accuracy. This implies that given any set of features, the multi-

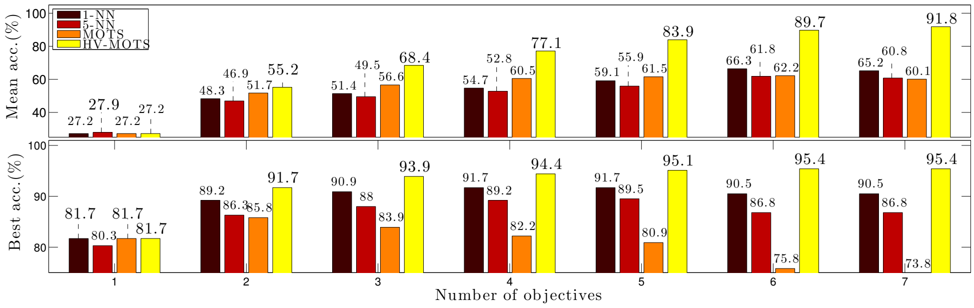



Figure 57: Classification results on the MuscleFish dataset for a growing number of objectives. For a given number of objectives, the left column indicates the mean classification accuracy and the right column indicates the best classification accuracy.

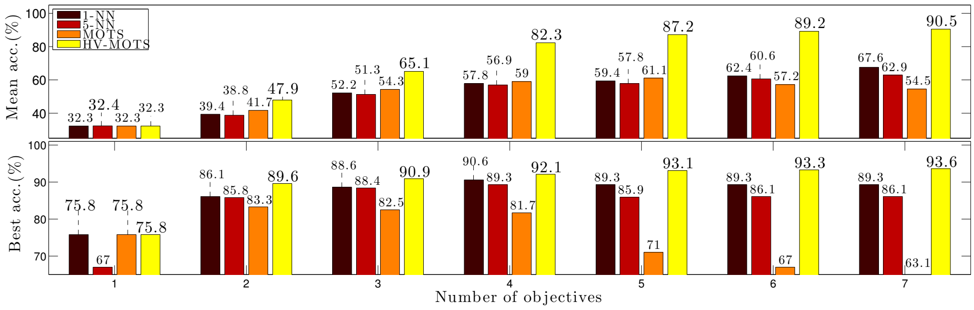



Figure 58: Classification results on the Freesound dataset for a growing number of objectives.

objective approach will obtain better results. To support this claim, we provide in Figure 59 the results of statistical significance tests between methods and across datasets, to rule out the effect of a particular data distribution. We follow the same guidelines as previously and provide the Tukey-Kramer HSD test [94] over the results of Friedman's ANOVA. We also present the statistical mean difference in accuracy over every combinations from a one-way ANOVA. Finally, we present the critical difference graphs which summarizes the column ranking of all methods over *every* features combinations. We can see in this figure that the mean column ranks and statistical mean difference in accuracy of HV-MOTS are strongly superior. The column rank corresponds here to the ranking of methods based on their accuracy results. This means that after two objectives, the HV-MOTS method is almost always in first position for any descriptor combination if ranked against other methods based on their accuracy score. Furthermore, the mean differences in accuracy increase with the number of objectives. It seems that the multi-objective classification is able to maintain the discriminative power of the best feature involved, whereas mono-objective selection will be confined by the worst features. This may go against the hypothesis that the feature set is more important than a particular learning scheme [249]. Furthermore, it seems here that the behavior of the whole class with respect to the input may be more important than the position of the input relative to the elements of the class. Therefore, even with lower dimensionality involved, the multi-objective paradigm is able to achieve a good classification accuracy. We can see that the MOTS paradigm (based on Pareto cardinality) is superior in mean classification accuracy to mono-objective selections for low dimensionality but starts to regress after four dimensions. This may come from the

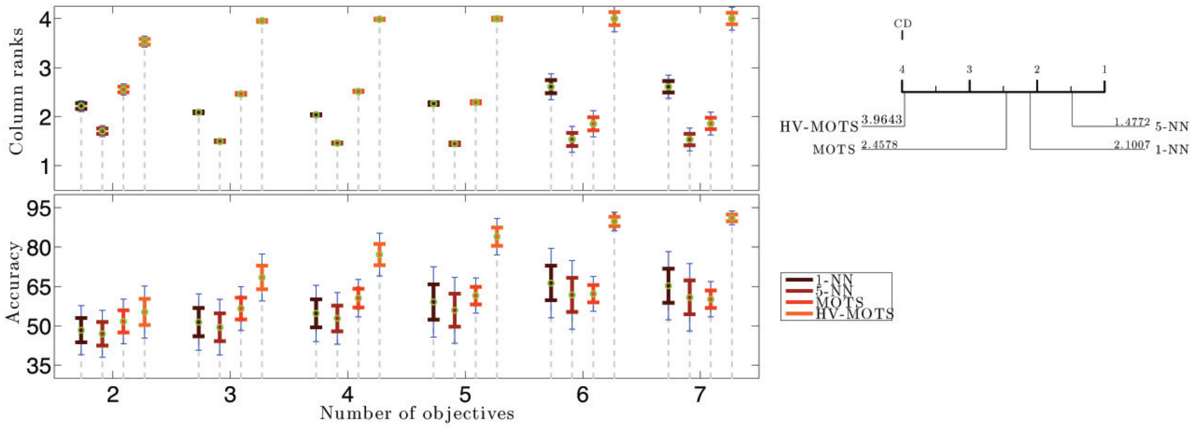



Figure 59: Significance tests between various methods for a growing number of objectives across both datasets. For a given number of objectives, the left column indicates the mean column rank (from Tukey-Kramer HSD over Friedman's ANOVA) and the right column gives the statistical mean difference in accuracy with the top performing method (from a one-way ANOVA).

fact that increasing number of dimensions creates more inclusive Pareto fronts which deludes the cardinality indicator. For other methods, performances stabilize and even regress slightly after five dimensions are involved.

We present in Table 15 the confusion matrix of the best classification accuracy (95.4%) obtained by HV-MOTS on the MuscleFish dataset. The corresponding descriptor combination is composed of MFCC, MFCCDeltaStdDev, PerceptualSlope, ChromaDeltaStdDev, RelativeSpecificLoudnessDeltaStdDev and PerceptualDecrease. It is interesting to note that most of the features used are related to the temporal behavior of the sound spectrum. The descriptors which are not temporal shapes are deviations of derivative, which in fact summarize the quantity of temporal variations for these descriptors. Furthermore, this combination contains descriptors for each structural aspects of sounds, namely energy (*Loudness*), harmony (*Chroma*), spectral shape (*MFCC*) and perceptual descriptors. It should be noted that the same accuracy was obtained by 18 similar combinations (which further confirms our intuition that HV-MOTS is able to retain the discriminative power of the best features involved). If we look at the distribution of the confusion matrix, we can outline different types of errors made by the system. First, the *class similarity* errors that can be expected when similar classes are part of datasets with widely diverse class types. For instance, elements of *male speech* are confused for *female speech* and the same apply to *violinbowed* confused with *cellobowed*. Second, the *morphological similarity* errors can be observed when the spectral behavior of two classes are alike. For instance, *violinpizz* are confused with *percussions* because of the impulsive nature of such sounds. The same applies to *machines* confused with *water* because of the long-term repetitive patterns that emerge from both. Finally, in both types can be found some *reciprocal errors* where the error applies symmetrically to two classes.

The HV-MOTS method was designed based on the hypotheses that temporal shapes would improve average features. At the same time, multi-objective selection is used to provide a perceptually more relevant and therefore more accurate classification. In order to analyze these hypotheses, we confront different views on experimental results. Figure 60 provides a comparison of the classification accuracy of using *only* temporal

| | Altotrombone | Animals | Bells | Cellobowed | Crowds | Laughter | Machines | Oboe | Percussion | Speech (female) | Speech (male) | Telephone | Tuba | Violinbowed | Violinpizz | Water |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Altotrombone | **13** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Animals | 0 | **8** | 0 | 0 | **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Bells | 0 | 0 | **6** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **1** | 0 |
| Cellobowed | 0 | 0 | 0 | **47** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Crowds | 0 | 0 | 0 | 0 | **4** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Laughter | 0 | 0 | 0 | 0 | 0 | **7** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Machines | 0 | 0 | 0 | 0 | 0 | 0 | **8** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **3** |
| Oboe | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **32** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Percussion | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **1** | **97** | 0 | 0 | 0 | 0 | 0 | **1** | 0 |
| Speech (female) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **35** | 0 | 0 | 0 | 0 | 0 | 0 |
| Speech (male) | 0 | 0 | 0 | 0 | 0 | 0 | **1** | 0 | 0 | **3** | **12** | 0 | 0 | 0 | 0 | **1** |
| Telephone | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **1** | 0 | 0 | **16** | 0 | 0 | 0 | 0 |
| Tublarbells | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **19** | 0 | 0 | 0 |
| Violinbowed | 0 | 0 | 0 | **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **44** | 0 | 0 |
| Violinpizz | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **2** | 0 | 0 | 0 | 0 | 0 | **38** | 0 |
| Water | 0 | 0 | 0 | 0 | **1** | 0 | **1** | 0 | 0 | **1** | 0 | 0 | 0 | 0 | 0 | **4** |

Table 15: Confusion matrix for the best classification accuracy (95.4%) obtained by HV-MOTS on the MuscleFish dataset. The descriptor combination used is composed of MFCC, MFCCDeltaStdDev, PerceptualSlope, ChromaDeltaStdDev, RelativeSpecificLoudness-DeltaStdDev and PerceptualDecrease.

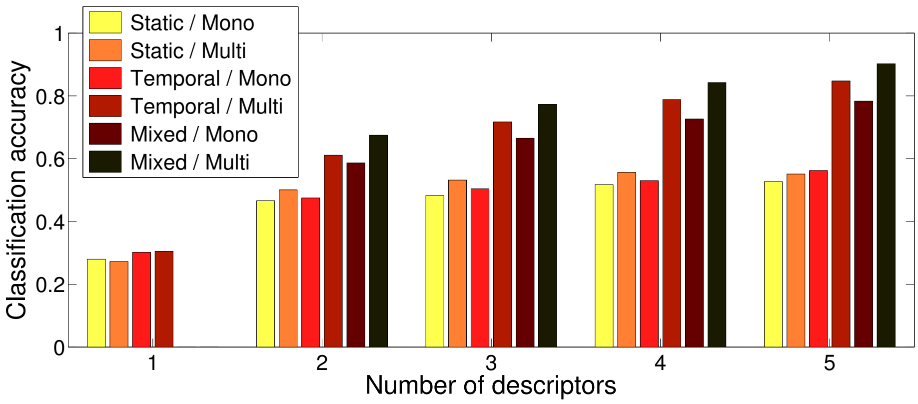



Figure 60: Comparison of the classification accuracy of using *only* temporal features, *only* static (mean and deviation) features or *mixed* sets of information with either multiobjective or mono-objective selection.

features, *only* static (mean and deviation) features or mixed sets of information. As we can see, the use of temporal features performs better than static features. More interestingly, it appears that best results are obtained by mixed sets of information, which indicates that normalized temporal shapes and static information are complementary sets of information. Finally, for any type of descriptors used, multiobjective selections performs consistently better than mono-objective approaches.

### 11.1.4 *Comparison to state of the art*

We compare our results to the state-of-art methods proposed with the same evaluation framework, namely a classification task on the MuscleFish dataset with a *Leave-One-Out* methodology. This allows to report published classification accuracies as a baseline for comparison. In their original study, Wold et al. [380] proposed to compute the mean, variance and autocorrelation of loudness, pitch, brightness and bandwidth, which together with duration amounts to a total of 13 features. By using a 1-NN rule, comparing the query to all feature vectors in the database with the Euclidean distance, they reported 80.9% classification accuracy. Guo et al. [154] later tested the applicability of a machine learning technique called *Boosting* based on a vector of 8 perceptual cepstral features which provided 78.3% accuracy. Guo and Li [153] proposed to use SVM on the same feature set and obtained 89% accuracy. Li [223] introduced the NFL method which was shown to provide 90.22% accuracy. Reyes-Gomes and Ellis [305] studied the use of GMM-EM and HMM with low entropy learning and obtained 89.9% accuracy. Finally, Shao et al. [325] used Neural Networks trained by GA with Back Propagation (BP-GA) over a set of 17 features and reported 92% classification accuracy. However their results are based on separate *Train* and *Test* sets procedure which does not allow straightforward comparison. Our HV-MOTS method allows to obtain a classification accuracy of 95.35% which outperforms previously reported accuracies for this dataset. The following table synthesizes the comparison between our method and previous approaches.

| | Proposed | |
|---|---|---|
| | Accuracy | N |
| Guo et al. [154] | 78.3 % | 8 |
| Wold et al. [380] | 80.9 % | 13 |
| Guo and Li [153] | 89.0 % | 8 |
| Reyes-Gomes [305] | 89.9 % | - |
| Li [223] | 90.2 % | 8 |
| HV-MOTS | **95.4 %** | **6** |

### 11.1.5 *Robustness analysis*

In real-life conditions, we can expect audio collections to include sounds from different sources recorded under various conditions. Some QBE systems have been tested for robustness but usually only with regards to transcoding, using either lower sampling rates [163] or lossy data compressions [67] to simulate mobile audio databases. We test our approach by applying a wider range of distortion classes to simulate various low-quality conditions in recording

- Additive white noise resulting in 30, 20 and 10dB SNR.
- Pitch down and upconversion by 10 and 20% of pitch.
- Random signal cropping by 5, 10 and 15% of length.
- Telephone filtering with a [300, 3400]Hz bandpass filter.

These distortions are applied one at a time to each sound clip. Modified samples are then used as queries to the database (minus the original non-distorted sample) which allows comparing *Leave-One-Out* classification accuracies after distortion. We use in these tests only combinations of the best feature sets obtained in the classification task with normal quality audio. Results of the robustness analysis are synthesized in Table 16. We can see here that HV-MOTS consistently outperforms other approaches for cropping, pitch modification and telephone filtering and appears to be robust for these transformations. However, it seems that both multi-objective approaches are more brittle than mono-objective selection when considering noise robustness. Although, it should be noted that the robustness of algorithms depends on the robustness of features.

## 11.2 SOUND MORPHOLOGY

The HV-MOTS classifier has already been put to use in a recent work in the domain of audio perception Koliopoulou [207]. The goal of this study was to investigate the notion of *sound morphology* introduced by the french composer Pierre Schaeffer Schaeffer [317]. He stated "*morphological profiles are meant to accurately describe the temporal evolution of some sound features and to propose an indexation and classification structure that could account for these evolutions*". We can see that the HV-MOTS classifier seems especially fit for this type of study.

| | Normal | Pitch conversion | | | | Telephone |
|---|---|---|---|---|---|---|
| | | -20% | -10% | +10% | +20% | |
| 1-NN | 91.69 | 88.02 | 90.71 | 89.73 | 86.06 | 90.95 |
| 5-NN | 89.24 | 85.09 | 87.29 | 87.04 | 83.37 | 88.26 |
| MOTS | 85.82 | 76.53 | 83.37 | 84.60 | 78.97 | 84.11 |
| HV-MOTS | **95.35** | **90.71** | **94.13** | **93.40** | **90.46** | **93.89** |
| | **Normal** | **Cropping** | | | **Noise (SNR)** | | |
| | | 5% | 10% | 15% | 10dB | 20dB | 30dB |
| 1-NN | 91.69 | 90.95 | 90.46 | 90.46 | **76.28** | 76.28 | 81.66 |
| 5-NN | 89.24 | 88.51 | 88.51 | 88.26 | 74.82 | 74.82 | 78.97 |
| MOTS | 85.82 | 84.60 | 84.11 | 84.11 | 66.26 | 66.26 | 74.08 |
| HV-MOTS | **95.35** | **94.87** | **94.13** | **93.89** | 74.82 | **78.97** | **84.84** |

Table 16: Effects of a set of distortions on classification accuracy for different methods on the MuscleFish dataset.

### 11.2.1 *Onset of the study*

This study comes within the scope of a wider research framework. This project tries to tackle the evaluation of interactive sound systems. The goals are to learn typologies of gestures associated with environmental sounds that could be exploited in *Sonic Interaction Design*. This field is directed at improving everyday interaction with tangible objects and interfaces. Its onset is therefore aimed at evidencing the morphological description of sounds. This requires categories of prototypical profiles for the sounds (classification), graphical representations (description) and temporal evolutions (dynamics, harmonic) of studied signals. The latest results in this topic Misdariis et al. [253] has evidenced the pertinency of this approach in the case of dynamic profiles. In this study, morphological classes have been formalized and associated with graphical prototypes

Therefore, a first perceptual experiment was carried out on a set of 55 environmental sound. First, this study was meant to define classes of morphological profiles (dynamic and melodic) adapted to these sounds. Then, the study aimed at conceiving a formalism to describe these profiles. Finally, the study tried to find a calculus from the temporal features in order to obtain the best signal representation for these classes. Therefore, the experiment was separated over two phases.

- Participants were asked to freely classify the sounds into 6 classes representing Schaeffer model, in order to regroup the sounds in each of the classes.
- Then, participants had to draw temporal profiles corresponding to these groups.

The first step allowed to obtain a coherent classification between different subjects by using a hierarchical clustering approach. The second part provided a set of morphological evolutions for each of the classes as presented in Figure 61.

The goal of the application of the HV-MOTS classifier for this problem was to evidence a coherence between the subjective clustering and the underlying audio features. The other aim of this study was to find the set of features that could best explain these morphological classes.

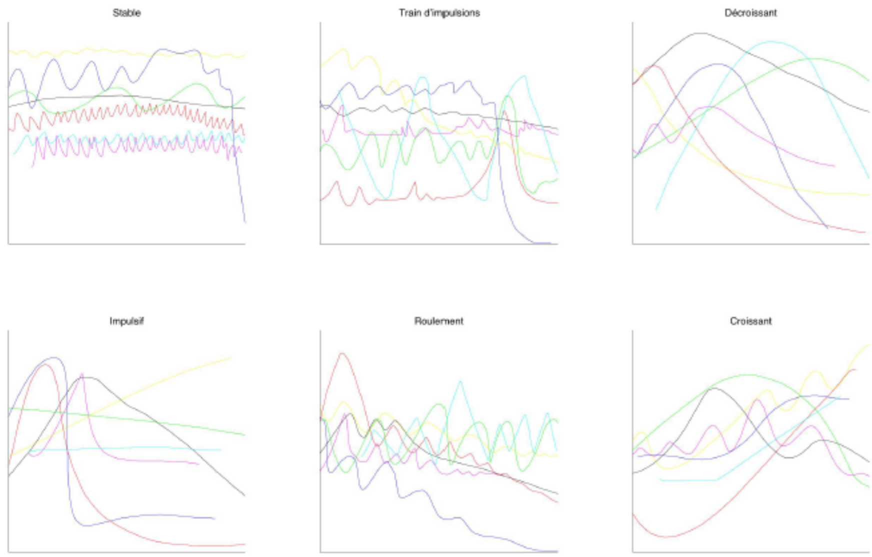



Figure 61: Temporal profiles drawn by each of the 19 participants for the set of 6 morphological classes.

### 11.2.2 *Results*

In order to exhibit this best set of audio features, the soundset is processed through the same workflow as the content-based audio application (cf. Section 11.1). Therefore each sound of the dataset is processed with IRCAMDescriptor in order to obtain every audio feature available (*Energy*, *Harmonic*, *Noise*, *Spectral* and *Perceptual*). The resulting time series are then normalized with *zero mean* and *unit variance* and resampled to a fixed length of 128 time points. Then, the classification of the subjective assessments is studied using the *Leave-One-Out* evaluation procedure for every combinations of audio features compared with DTW distance measure. The results of the classification for an increasing number of objectives is presented in Figure 62.

As we can see, the HV-MOTS classification strongly outperforms the other approaches. If we compare these results to those of the large scale study (cf. Section 9.3), it seems that the margin between 1-NN and HV-MOTS accuracies are even larger in this particular dataset. The maximal accuracy obtained at level 6 is *69.09*% for 1-NN as opposed to an accuracy of *87.27*% for HV-MOTS. This result correlates well with the hypothesis on which our proposal has been constructed. Indeed, the HV-MOTS approach was constructed to mimic our auditory perception, by producing a multi-dimensional assessment of temporal similarities. Therefore, its strong superiority on an audio perception problem further confirms the validity of these hypotheses. The second goal of this study was to obtain the set of audio features that could best explain the subjective classification. Table 17 presents the best features combination obtained by the classification step.

As we can see, the best combination embeds several aspects of the sound properties. Two features describe the *energy* components of sound (*energy envelope* and *loudness delta*). The *fundamental frequency* allows to describe the evolution of *pitch*. For the *harmonic* content, the inharmonicity describes the evolution of harmonic peaks while the *noisiness* gives the ratio between the noise and harmonic components of a sound.

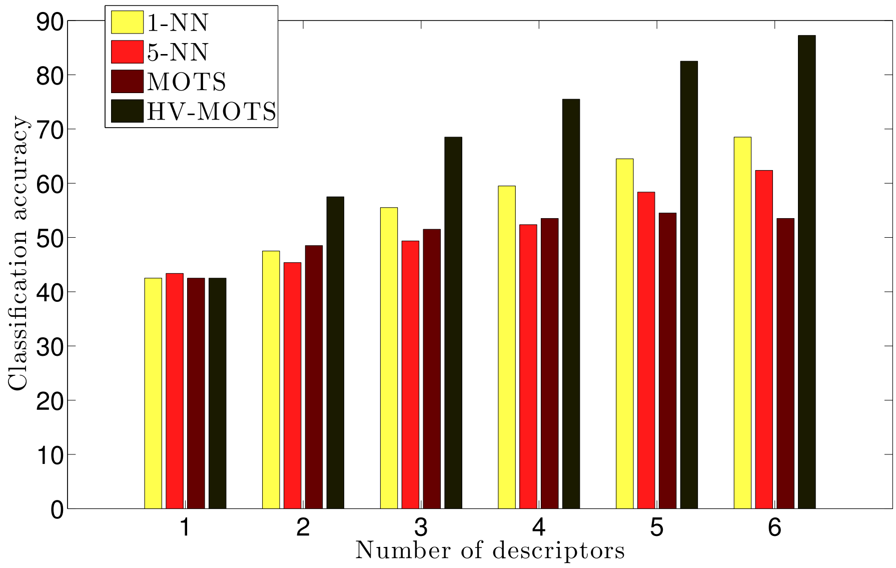



Figure 62: Accuracy of different classification methods for an increasing number of descriptors over the sound morphology.

| **Best combination obtained** | **(Accuracy : 87.27%)** |
|---|---|
| *Energy Envelope* | 42.51 % |
| *Spectral Rolloff* | 37.82 % |
| *Loudness Delta* | 28.83 % |
| *Noisiness* | 21.13 % |
| *Inharmonicity* | 19.24 % |
| *Fundamental Frequency* | 18.65 % |

Table 17: The best combination (6 features) obtained thanks to the HV-MOTS classification paradigm provides a classification accuracy of 87.3%. The right column shows the individual classification accuracy for each feature.

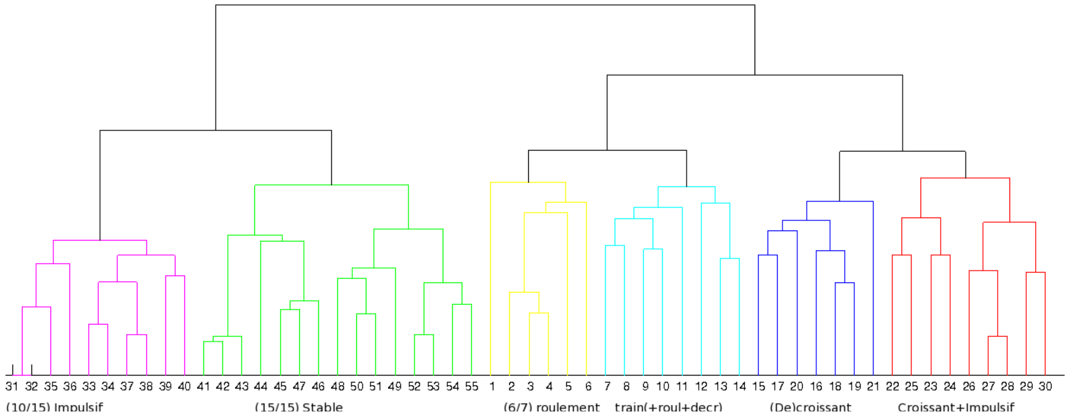


Figure 63: Results of a hierarchical clustering performed on the sounds of the study using the audio features selected thanks to the HV-MOTS classification paradigm.

Finally the *spectral rolloff* provides a description of the overall position of the energy in the spectrum. To validate this set of features, Figure 63 presents the results of the hierarchical clustering performed on the sounds of the study using the selected audio features. The clustering is performed by computing the cityblock ($L_1$) distance between pairs of features and then the shortest distance between pairs is used to regroup sounds in clusters.

The selected features allow to obtain almost the same clustering as obtained from the subjective ratings. As we can see in this figure, three of the classes are perfectly clustered. For the other classes, almost all sounds have been clustered in their corresponding set. Therefore, the HV-MOTS approach allowed to obtain a strongly higher classification with a set of features which is consistent with perceptual studies. The remaining part of this study Koliopoulou [207] exhibits that this set allows to compute a meaningful set of prototypical profiles for each morphological class. We do not include this part for the sake of clarity and redirects the interested readers to Koliopoulou [207].

# Part V

# GOING BACK TO MUSIC

# 12 ORCHESTRATION

We now get back to our original artistic problematic. As we discussed in Chapter 1, musical orchestration is an art of musical writing that rely on the spectral characteristics unique to each instrument. Therefore, it goes beyond the realm of symbolic processing and thrives in the territory of signal possibilities. Trying to tackle the problem of orchestration from a scientific angle involves the use of unformalized knowledge and unveils numerous facets of complexity.

## 12.1 ON THE COMPLEXITY OF ORCHESTRATION

The history of music is littered with critical moments when the established conventions are no longer sufficient to accompany the novel musical trends. Therefore, when scientific research focuses on musical issues, there is no choice but to follow this need for complexity. We will try to highlight some open problems in different areas of complexity in the orchestration topic. We divide this discussion into three major areas of complexity: *combinatorial* complexity, *temporal* complexity and *timbre mixtures* complexity.

### 12.1.1 *Combinatorial complexity*

The first facet of orchestral complexity arise from the exponential number of possible instrumental mixtures. Indeed an orchestra can count up to hundreds of players and is usually composed of a large-scale assortment of instrumental groups. Each instrument is able to produce a variety of playing modes on a vast range of notes which can be played at various intensities. Therefore, trying to find every combination of timbre that can be played by an orchestra implies to solve a NP-Complete problem. Furthermore, it is hard to predict the properties of every instrumental mixture, as it is computationally intense to compute the signal features on each mixture. We can take a quick sobering experiment to illustrate this problem. If we suppose that each instrument can play only 25 notes (a grand piano can play up to 96 notes) at 2 dynamics (the symbolic notation contains 9 dynamic symbols but crescendos can produce continuous variations) with only 2 playing modes. This yields $10^2$ possible musical atoms for each instrument. Hence, a set of $N$ instruments would lead to $10^{2N}$ possibilities. Now, if we suppose that our computer is able to evaluate the spectral features of a combination in only $10^{-9}s$ ($1ns$ is extremely fast considering that this computation is currently at most real-time), then evaluating all combinations requires

$N = 5$ $\quad$ $10^{10}$ combinations $\rightarrow$ 10 seconds

$N = 8$ $\quad$ $10^{16}$ combinations $\rightarrow$ 115 days

$N = 10$ $\quad$ $10^{20}$ combinations $\rightarrow$ 3.170.979.198.376 years

This small experiment allows to get a glimpse on the extent of this problem. Furthermore, we have taken here extremely simplifying assumptions. It is therefore mandatory to find an approach that can handle this combinatorial explosion.

### 12.1.2 *Temporal complexity*

As we discussed in Section 2.4, the temporal structures are of prime importance in music processing. However, if we look back at Figure 1, reaching back to the scale of orchestration reveals a new dimension of complexity. Indeed, the *macro-temporal* evolutions combine vertical writing (the arrangement of different voices in a restricted time frame) and horizontal writing (the development of the complete musical structures over time). Therefore, orchestration unveils an interaction between the *micro-temporal* properties of musical atoms and the *macro-temporal* articulations of their organization. Each scale relates to different levels of complexity. We have studied the micro-level extensively throughout this document. However, the macro-level is a necessary condition for the orchestration of polyphonic sequences. As we will discuss in Section 12.2.2, we address the vertical and horizontal dimensions of orchestration simultaneously. Therefore, we must face the problem of macro-temporal timbre which evolves continuously. It seems impossible to consider this dynamic orchestration as a mere extension of a static model in where the notion of time would come down to a series of segmented instants. Hopefully, this is where all the knowledge gained in previous chapters will prove its usefulness.

### 12.1.3 *Orchestral timbre*

The advent of computer music has allowed to explore a novel universe of sound possibilities. The electronic instruments, introduced "unbelievable" timbre, profiling a new dimension in the imaginary of sound. These are now an integral part of contemporary musical discourse. The timbre gradually became a central element of this new musical language and is now even looked upon as the backbone of musical writing. The large number of instruments and their variety of timbre offer a virtually infinite palette of orchestral colors. However, as we discussed in Chapter 1, there are numerous problems when trying to cope with the multidimensional aspect of timbre and the prediction of sound mixtures properties. One of the main facet of complexity in this context is to understand the coupling between instruments that arise from sound mixtures.

THE PHENOMENA OF EMERGENCE These phenomena are still poorly understood today but are embedded in the study of orchestration. Indeed, it is well-established that two instruments create a timbre different from the simple sum of their respective timbre. It is the relationship between instruments that create different timbres, because of their complex acoustic interactions. Therefore, these phenomena can cause the appearance of elements in the spectrum of a mixture that are not in any of its constituents. Reciprocally, the *masking* and *phase* effects can cause the disappearence of components because of their relative levels or closeness of spectrum (which illustrates the idea that the mixture spectrum does not follow a linear addition). The phase effect has been extensively studied and also seems deeply involved in recognition of the instruments. Other interesting phenomena can emerge from an accumulation of sound sources such as the phenomena of *unison* and *chorus*. Conversely, the instrumental *fusion* can create the illusion that two instruments merge their sounds to the point that they are impossible to discriminate.

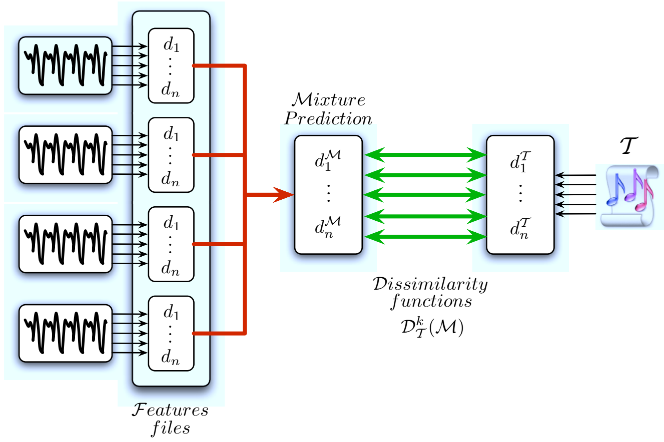



Figure 64: Original approaches to tackle the problem of computer-aided orchestration. A target sound file is fed to the system which will try to reconstruct it by using a set of instruments. The system rely on a set of feature files which are combined through prediction functions.

## 12.2 HOW TO REIFY ORCHESTRATION

After seeing these facets of complexity, we could wonder how to approach the orchestration on a scientific basis. Automatic orchestration is a very recent topic in computer-aided composition, for which only few systems exist today. In order to find a single point of entry to this problem, the systems all rely on the concept of a *target* to be optimized. An orchestration problem then consists in finding a mixture of instrument samples that minimizes different spectral distances. This approach is illustrated in Figure 64, where the goal is to find the mixtures of instruments that best match a given timbre. The user specifies the instruments he would like to use (*constraints*) and the characteristics of the sound to be produced (a sound *target*). This target is a sound file which defines the audio features to reproduce. Then, an orchestration engine relies on an instrumental *features database* to suggest instrument combinations (orchestration proposals) that sound close to the target. The problem is therefore to find an efficient way to converge towards closely sounding elements (and hence circumvent the combinatorial complexity).

### 12.2.1 *Existing systems*

Psenicka [290] first developed a system called SPORCH (SPectral ORCHestration) where the search is based on an iterative matching pursuit algorithm. Each instrument in the database is associated with a series of pitch, dynamics and a collection of the perceptually most significant harmonics. The target is analyzed with a spectral analysis procedur. The algorithm first selects the best matching instrumental sample. The spectrum of this sample is then subtracted from the target and removed from the database. The algorithm iterates while trying to minimize the Euclidean distance between the target and the current mixture. Rose and Hetrick [309] later proposed a tool to analyze existing orchestrations and propose new ones. The instrumental knowledge is summarized by the average harmonic spectrum calculated on the sustained part of the instruments. The algorithm is based on a Singular Value Decomposition (SVD) of the spectrum contributions. The target is finally expressed as a weighted sum of the spectra contained in the database. Another system was subsequently proposed by Hummel [174]. An iterative algorithm is also used, but the distance is based on the spectral envelope rather than the harmonic partials. The system calculates the spectral envelope of the target and repeatedly search for the best approximation. Consequently, the harmonic structures can be very different between the target and the resulting mixture. The author advise to use his system with sounds lacking pitch such as whispered vowels.

Recently, an interesting approach was proposed by Carpentier [71]. Although the formulation of the problem remains the same [72], the system is based on multiobjective genetic exploration [352]. This allows to propose a set of optimal solutions, rather than a single instrumental combination. The target is represented by a set of audio features that describe different aspects of the sound to be reproduced with a pre-determined orchestra. Using a genetic algorithm, the system retrieves the Pareto front of solutions. The user then selects a solution, which allows the system to infer its preferences among the different dimensions. The instrumental knowledge is based on the model developed by Tardieu [351]. This model uses Gaussian Mixture Models (GMM) to learn the distribution of features for a large number of samples and allows to infer the properties of missing ones. Furthermore, a set of functions [350] were developed for each feature which can predict the characteristics of a mixture of instrumental sounds.

### 12.2.2 *Discussion*

We try to highlight here the shortcomings of existing orchestration systems.

- The main goal of every system is to approximate a *sound target*. This restricts their scope of study to a very narrow case of orchestration. Indeed, most of the time, composers does not have a well-formed example of the sound to obtain. The orchestration generally aims to produce, not to re-produce a timbre.

- These systems make the implicit assumption of linear additivity of the timbres. However, the simple consideration of the phase effect is enough to put it into default. The predictive capacity of an additive model may be doubted, despite its computational advantages.

- Finally, the most important shortcoming of every previous system is that they only provide vertical orchestration by focusing the analysis on sustained instruments. Therefore, even the best proposed solutions completely neglect the temporal evolution of the sound target. The instrumental knowledge is, therefore, limited to sustained harmonic sounds without any temporal variations.

As all previous systems rely on "time-blind" features, they are only able to provide static orchestrations. However, as we discussed in Section 2.4 the territory of timbre is not confined to a static structure of proportions. It rather comprises "variation laws" in a context continuously evolving over time. It is therefore essential to move to a higher level of modeling, by assessing the complete temporal structure of spectral features. Advantages of this approach are twofold. First, the generated orchestrations are improved and more realistic by reproducing the whole spectro-temporal structure. Second, it allows the use of evolving playing modes like crescendo, glissando, multiphonics and so on. For all these reasons, we focus on providing a system that addresses the problem of time in musical orchestration. However, we will also tackle the two

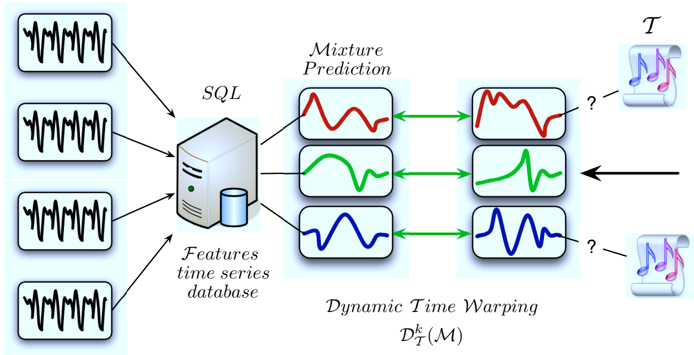



Figure 65: A new approach for computer-aided orchestration. Unlike previous systems (cf. Figure 64), the comparison between targets and sound mixture is based on their temporal evolutions. The knowledge is encapsulated in a SQL database in order to provide an almost infinite source of knowledge. Finally, the need of a well-formed audio example is bypassed by the direct input of temporal shapes.

other aforementioned shortcomings. Hence, we introduce the notion of *abstract target* and also bypass the additive model by using *prediction functions.*

## 12.3 GOING FURTHER IN COMPUTER-AIDED ORCHESTRATION

We outline here our main goals when devising an orchestration system that could take into account the temporal evolution of audio features. Unlike previous systems (cf. Figure 64), the comparison between the target and sound mixtures is based on the *temporal evolution* of their audio features. The knowledge is encapsulated inside a SQL database in order to provide an almost *infinite source of knowledge* (without loading all features in live memory). Finally, the need of a well-formed audio example could be bypassed by providing *abstract targets* such as the direct input of the temporal shapes. The final workflow embedding these improvements is shown in Figure 65.

### 12.3.1 *Algorithmic choices*

*Knowledge database*

The previous systems all rely on a set of files that contains the static audio features of instrumental sounds. This source of knowledge is inherently limited as it requires to load the complete database in live memory (in order to avoid the cost of disk accesses along the search algorithms). Therefore, the systems must limit the quantity of their own knowledge. This memory-based approach seems to be wildly condradictory with the combinatorial complexity that we exhibited in Section 12.1.1. Furthermore, as we intend to focus on temporal shapes, the dimensionnality of data to be processed will be greatly extended. Hence, we decided to use a SQL database architecture in order to store the audio features. The advantages of this choice are two-fold. First, it allows a dynamic and almost infinite source of knowledge. Second, by using the work previously presented (cf. Section 11.1), this database provides the MOTS, MOSEQ and QVI embedded inside the orchestral knowledge. We will see that these improvements can greatly enhance the orchestration search algorithm. (cf. Section 12.4.2).

*Temporal matching*

This is where all the knowledge gained from the past chapters can be put to use to improve our original artistic problematic. First, the use of the MOTS framework (cf. Chapter 7) in the database allows to perform temporal queries on the instrumental knowledge as well as using the MOSEQ and QVI paradigms (cf. Section 7.5). Then, when trying to optimize the sound mixtures to match a given target, we assess their temporal similarity with the *Dynamic Time Warping* (DTW) distance measure. Hence for each sound mixture, we forecast its audio features by using the prediction functions provided by Tardieu and Rodet [351]. Even if these functions were devised for static values, we consider that they can be extended to time series on a point-by-point basis. Based on these predicted time series, we consider the DTW as a dissimilarity measure to select the best mixtures.

*Abstract target*

The concept of a *sound* target restricts the scope of study to a very narrow case where composershave a well-formed example of sound to obtain. However, since we provide a new approach that can handle temporal matching, we can bypass this simplification. As presented in Figure 65, we provide several forms of targets ranging from sound files to purely abstract targets. First, the user can still input a single sound file to the system. The analysis module then show the temporal shapes of the corresponding sound, which the user is free to modify. Then, the system offers the possibility to use *multiple* sound files, by selecting the temporal shape of a first target to be combined with the feature of other sound files. The user can input temporal shapes independently of the sound files. Finally, a *purely abstract* target can be created in which all the features to optimize are the result of a direct input procedure.

## 12.4 ABSTRACT TEMPORAL ORCHESTRATION - MODULAR STRUCTURE (ATO-MS)

We introduce in this section a system for temporal orchestration generation that address the shortcomings of previous systems. *ATO-MS* is an optimization system which allows to find sound combinations that approximate the temporal evolution of several audio features. By relying on multiobjective genetic algorithms, we can evolve *populations* of sound mixtures that jointly minimize a set of objective functions. These functions are defined as the distance between selected audio features. Therefore, each distance can either represent the DTW between time series features, or the Euclidean distance between average features. Each *individual* in the population represents a potential sound mixture. The individuals contains the symbolic properties of the mixtures, encoded inside a *genome*. Therefore, each genome defines an orchestral score which will be used to approximate the target. We define the orchestral genome in the following way

| $(\mathtt{Instrument}_1)$ | | | $\dots$ | $(\mathtt{Instrument}_N)$ | | |
|---|---|---|---|---|---|---|
| $\mathtt{Sample}_1^{id}$ | $\mathtt{Onset}_1$ | $\mathtt{Duration}_1$ | $\dots$ | $\mathtt{Sample}_N^{id}$ | $\mathtt{Onset}_N$ | $\mathtt{Duration}_N$ |
| $[0 \dots \vert DB \vert]$ | $[0 \dots \vert \mathcal{T} \vert]$ | $[0 \dots \vert \mathcal{T} - O_1 \vert]$ | $\dots$ | $[0 \dots \vert DB \vert]$ | $[0 \dots \vert \mathcal{T} \vert]$ | $[0 \dots \vert \mathcal{T} - O_N \vert]$ |

The set of instruments to be used by the system is defined by the user prior to the search. For each instrument, the $\mathtt{Sample}_i^{id}$ defines the index of the sample to use in the database. This sample implicitly defines the note, playing mode and dynamics of the

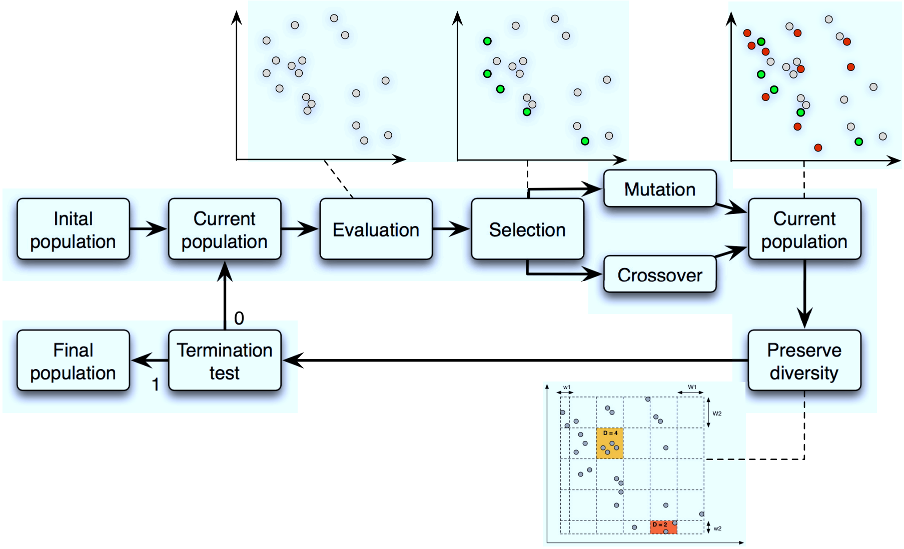



Figure 66: Algorithmic workflow of the *multiobjective genetic algorithm* for orchestration generation. We display an example population and how the different steps can affect it.

corresponding instrument. $\mathtt{Onset}_i$ defines the starting time of the $i^{th}$ instrument and $\mathtt{Duration}_i$ its playing length. Hence, the goal of the genetic algorithm is to make these individuals evolve by trying to mimick the process of natural selection, as illustrated in Figure 66.

The algorithm starts with an initial population, which is usually randomly selected (we will show that we can enhance the performances by providing a more elaborate seeding). Then, the population is evaluated, which implies in our context, two complementary computations. First, for each individual, the audio features of the corresponding mixture are predicted with the functions provided by Tardieu and Rodet [351]. Then, the similarity between every individual and the target is computed along each objective. Therefore, the temporal evolutions of audio features are compared with the DTW measure, and the average audio features are compared with the Euclidean distance. Then, the efficient solutions inside the population are extracted using the Pareto optimality. This is where the imitation of genetic evolution is being applied through the use of genetic *operators*. First, the individuals in the population are being mated randomly. From their union, results an *offspring* which share the genomes of their parents obtained through a *crossover* procedure. We use the *1-point crossover*, where all genes before a random point come from the first parent and all genes after this point come from the other. Finally, the *mutation* operator changes randomly selected genes from individuals in the pool. When all the operations have been performed, the diversity of the population is preserved using the PADE procedure [71]. This means that when the population exceeds a threshold, the individuals to be removed are selected from the most crowded regions.

This algorithmic workflow is the traditionnal application of multiobjective genetic algorithms and the exact procedure followed by the previous orchestration system [71]. We could apply this method to solve our temporal orchestration problem by simply using the DTW as an objective function. However, the use of time series data strongly increases the combinatorial possibilities (and therefore complexity) of the problem. Therefore, we show in the next sections how to provide more elaborate optimizations

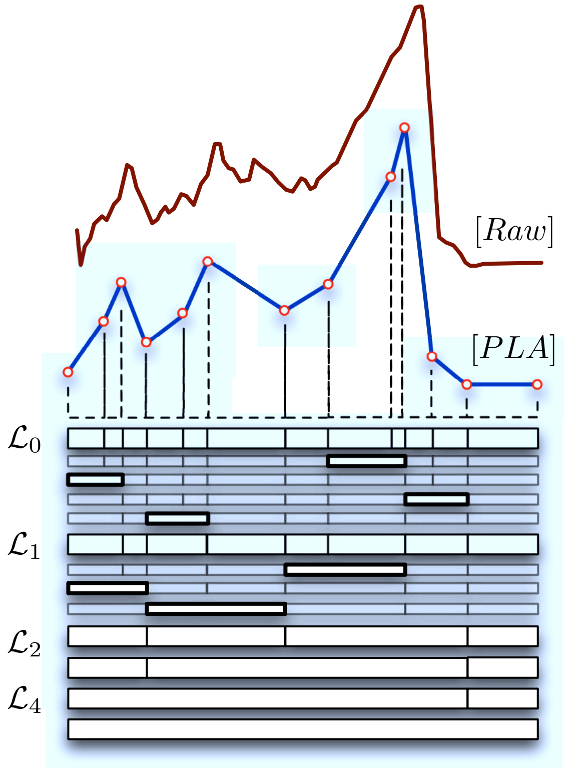



Figure 67: Illustration of the *multi-level entropic segmentation* procedure. From the raw time series, we first extract a PLA representation by using a bottom-up algorithm. Then, we start to regroup these segments based on the variation of entropy they provide. Each time a segment previously merged is going to be merged again, the current level is increased.

to enhance the results. First, we introduce a segmentation procedure specific to our problem, that will allow to detect significant segments in the temporal evolutions. This will allow to divide the problem by finding coherent events inside the targets. Then, we show how to combine this segmentation with a local optimization in order to enhance the solutions proposed by the algorithm.

### 12.4.1 *Entropic segmentation procedure*

We introduce here a novel segmentation procedure that can account for various levels of information. In our context, the main drawback of segmentation procedures is that they usually seek to define the *onsets* of events (*musical atoms*). However, we need to "break down" further this musical unit as we seek to reconstruct it with a mixture of instruments. Therefore, we need a multi-level segmentation that can give access to parts of the time series that exhibit a sense of coherence. For that purpose, we perform a segmentation based on the *approximate (Shannon) entropy*. This provides a measure of the average information content in each segment. Our segmentation method first relies on a *bottom-up* Piecewise Linear Approximation (PLA) (cf. Section 5.4.2). Then, we use this representation and regroup the remaining segments based on the variation of entropy they provide. These ideas are illustrated in Figure 67.

Hence, we start by considering each pair of time points as a potential coherent segment. We then work by agglomerating different segments together. At each iteration,

**Algorithm 12.1** The *multi-level entropic segmentation* procedure

```
multiLevelSegmentation(data, thresh, minSegs)
 // Perform Piecewise Linear Approximation
 data ← (data−μ_d)/σ_d
 segments ← [1 ... N_data]
 // Initialize the cost of merging each pair of segments
 for i ∈ [1 ... N_seg − 1]
   cost_i ← error(data, merge(segment_i, segments_{i+1}))
 end
 // Keep merging until reconstruction error or number of segments is attained
 while errRecon < thresh || N_seg > minSegs
   id ← min(cost)
   segments_id ← merge(segments_id, segments_(id+1))
   remove(segments_(id+1))
   cost_(id−1) ← error(data, merge(segments_(id−1), segments_id))
   cost_id ← error(data, merge(segments_id, segments_(id+1)))
   errRecon ← D_L2 (data, segments)
 end
 // Initialize the variation of entropy induced by pairs of segments
 for i ∈ [1 ... N_seg − 1]
   Δ^i_entropy ← entropy(merge(i, i + 1)) − (entropy(i) + entropy(i + 1))
 end
 // Start the multi-level entropic grouping
 while N_seg > 1
   id ← max (Δ_entropy)
   segments_id ← merge(segments_id, segments_(id+1))
   // If one of the segments has already been merged in this level, move one up
   if hasBeenMerged(id) | hasBeenMerged(id + 1)
     finalSegments_level ← segments
     level + +
   end
   remove (segments_(id+1))
   update (Δ_entropy)
 end
```

we select the merge operation that induces the minimal reconstruction error. This procedure is repeated until the reconstruction error is above a threshold or a minimum number of segments is reached. This gives us a first segmentation based on the PLA representation. Then, we perform the *multi-level* grouping on this list of segment. Hence, we compute for each pair of segments the variation of entropy they induce. Then, we operate the same grouping as before, but this time we select the maximal entropy variation (we try to regroup segments that contain the most information together). Each time a segment previously merged at this level is going to be merged again, the current level is increased. We present in Algorithm 15.1 the final implementation of the *multi-level entropic segmentation* procedure.

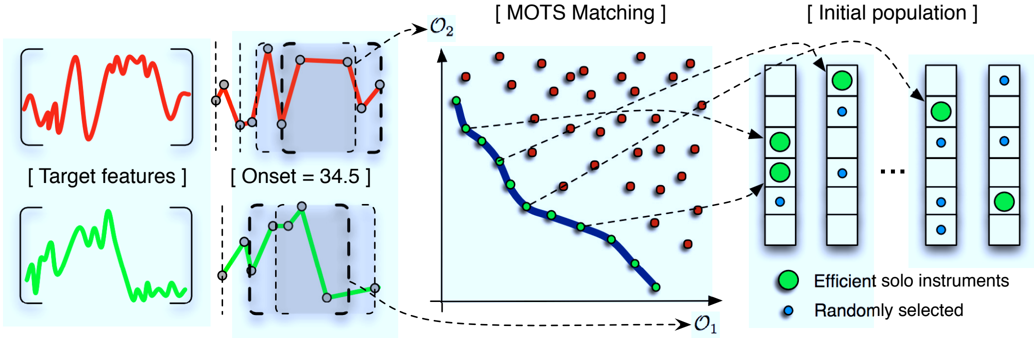



Figure 68: We combine paradigms by using the MOTS algorithm and the entropic segmentation in order to find efficient solo instruments. These can be used as "seeds" to the initial population. That way we reduce the blindness of pure randomness

### 12.4.2 *Optimal warping*

We now introduce the algorithm for computer-aided temporal orchestration. Overall this proposal can be classified as a multi-objective genetic optimization algorithm. However, we enhance this approach by using a specifically designed *local optimization* method. This procedure strongly rely on the MOTS paradigm and the entropic segmentation in order to refine the proposed solutions after each iteration of the algorithm. We also use this strategy for generating the initial population.

MOTS - INITIAL SEEDING One of the well-known problems in genetic algorithms is that the final solutions might be strongly influenced by the initial population. Furthermore, most of these algorithms usually rely on a blind selection for the initial population generation. Hence, individuals are being chosen randomly prior to the search procedure. Because of the increased complexity of time series, this blind selection might be insufficient to provide closely matching solutions. Therefore we combine the entropic segmentation procedure with the MOTS matching algorithm as a "kick-start" for the multiobjective genetic algorithm. This idea is depicted in Figure 68.

We start by performing an entropic segmentation on each of the target time series. This leads to a set of segments for each series. Then, for each of the segments, we select the corresponding data in *every* time series. Therefore, for a given segment in one feature, the corresponding subsequences are extracted from the others. These subsequences are then used as inputs to the MOTS matching algorithm. The efficient results from the Pareto front are then added to a *seeding pool* for the initial population, with corresponding onsets and durations. This leads to a final *constrained random selection* of the initial population (some of the elements are still selected by random decision). That way, the blindness of the initial population is reduced, as several efficient instruments are included prior to the search.

MOTS - LOCAL OPTIMIZATION The previous procedure allows to reduce the blindness of the initial population selection. However, after the search procedure is initiated, there is no control over the evolution of solutions. As we discussed earlier, because of the complexity of time series data, we might need more elaborate controls. Therefore, we introduce an analogous local optimization procedure after each iteration of the algorithm. This procedure is aimed at selecting some of the already efficient solutions

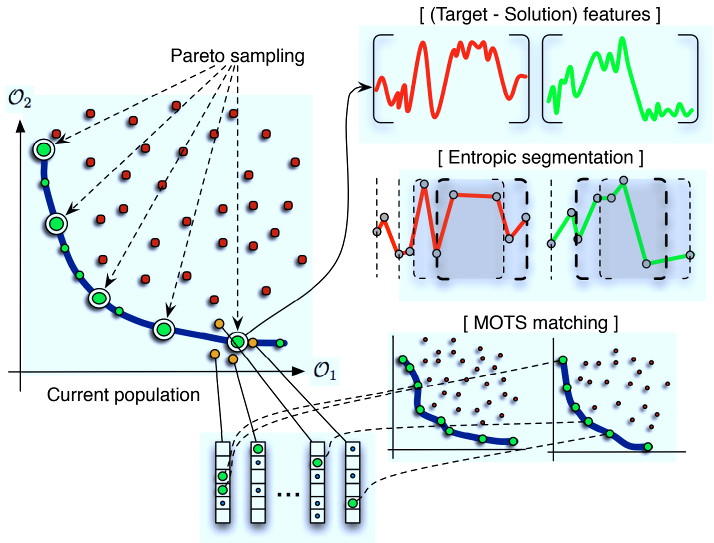



Figure 69: We apply the MOTS algorithm further in order to provide a local optimization loop after each iteration of the algorithm. This procedure allows to enhance each of the selected solution obtained by sampling the Pareto front.

and then optimizing them individually. Therefore, its goal is to provide a new set of more efficient solutions to include in the population. These ideas are depicted in Figure 68.

As we can see, we start by selecting solutions in the current Pareto front. This is handled by sampling the front, ie. we select a fixed number of solutions inside the current set. Then, for each solution, we compute the differences between its current features and the target features. This leads to a new set of time series. Then, we perform the same procedure as the initial seeding. First, we perform an entropic segmentation on each time series and use these subsequences as inputs to the MOTS matching algorithm. Then, we generate *new* individuals that are produced by inserting the efficient results of the MOTS matching inside the original solution. Finally, the resulting individuals are evaluated and included in the current population.

ALGORITHM We now introduce the complete *Optimal Warping* algorithm for solving the problem of temporal orchestration, which is detailed in Algorithm 15.2. We start by generating a first constrained population by using the initial seeding procedure presented previously. This population is first evaluated (the features of each individual are predicted and compared to the features of the target). Then, the Pareto front is extracted from the set of initial individuals. Then, the multiobjective genetic search loop is processed. At each iteration of the search, we select a portion of the population (the genetic *pool*) that will be used to perform genetic operations. Therefore, individuals in the pool are mated (*crossover*) and some are *mutated*. This is where we apply our *local optimization* procedure. First, efficient solutions are selected by sampling the Pareto front. Then, for each of these solutions, we compute the difference between their features and the target. The resulting time series are used to perform a segmentation. The MOTS algorithm provides a set of efficient solutions which are used to generate new individuals. The genetic operators and local optimization generate new sets of individuals which are evaluated and added to the current population. Then, the Pareto

**Algorithm 12.2** Our proposed *Optimal Warping* algorithm which combines a multi-objective genetic algorithm and a local optimization procedure, based on a combination of the MOTS framework and the entropic segmentation method.

**optimalWarping**($\mathcal{T}$)
// Perform initial seeding procedure
$\text{segments} \leftarrow \text{multiLevelSegmentation}(\mathcal{F}^k(\mathcal{T}))$
$[\text{seeds}, \text{onsets}] \leftarrow \text{MOTS}_{match}(\text{segments})$
$\text{pop} \leftarrow \text{constrainedRandomPopulation}\left(N_{pop}^{init}, \text{seeds}, \text{onsets}\right)$
$\text{criteria} \leftarrow \text{evaluatePopulation}(\text{pop}, \mathcal{T})$
$\text{pareto} \leftarrow \text{extractPareto}(\text{criteria}, \text{pop})$
**while** $\text{iter} < \text{iter}_{max}$
  $\text{pool} \leftarrow \text{selectTournament}\left(\text{pop}, N_{pool}\right)$
  $\text{pool} \leftarrow \text{crossover}(\text{pool})$
  $\text{pool} \leftarrow \text{mutation}(\text{pool})$
  $\text{pop} \leftarrow \text{pop} \cup \text{pool}$
  // Perform the local optimization procedure
  $\text{select} \leftarrow \text{sample}\left(\text{pareto}, N_{optim}\right)$
  **for** $i \in \left[1 \ldots N_{optim}\right]$
    $\text{segments} \leftarrow \text{multiLevelSegmentation}(\mathcal{F}^k(\mathcal{T}) - \mathcal{F}^k(\text{select}_i))$
    $[\text{seeds}, \text{onsets}] \leftarrow \text{MOTS}_{match}(\text{segments})$
    $\text{optimized} \leftarrow \text{generateIndividuals}(\text{select}_i, \text{seeds}, \text{onsets})$
    $\text{pop} \leftarrow \text{pop} \cup \text{optimized}$
  **end**
  $\text{criteria} \leftarrow \text{evaluatePopulation}(\text{pop})$
  $\text{pareto} \leftarrow \text{updatePareto}(\text{criteria}, \text{pareto})$
  **if** $N_{pareto} > N_{pareto}^{max}$
    $\text{pareto} \leftarrow \text{PADE}_{decrease}(\text{pareto}, N_{pareto}^{max})$
  **if** $N_{pop} > N_{pop}^{max}$
    $P \leftarrow \text{pop} \backslash \text{pareto}$
    $P \leftarrow \text{PADE}_{decrease}(P, N_{pareto}^{max})$
    $\text{pop} \leftarrow P \cup \text{pareto}$
  **end**
**end**
**return** pareto

front is updated accordingly. Finally, if the Pareto front or the population exceeds their fixed maximum, their cardinality is decreased by using a diversity preservation procedure.

### 12.4.3 *Comparison with Orchidee*

In order to evaluate the proposed improvements, we will compare them to the previous best performing algorithm as a baseline. In order to perform a fair and objective evaluation, we use exactly the same evaluation procedure proposed in Carpentier [71]. Therefore, the algorithms are evaluated on 500 monophonic problems with cardinality and pitch constraints, 500 polyphonic problems with same constraints, 500 unconstrained monophonic problems and 500 unconstrained polyphonic problems. All the algorithms use the exact same set of parameters which are $N_{pop}^{init} = 200$, $N_{pop}^{max} = 500$, $N_{mate} = 50$, $N_{pareto}^{max} = 200$, $\text{Iter}_{max} = 100$.

| | Monophonic constrained | | Monophonic unconstrained | |
|---|---|---|---|---|
| | Orchidée | Optimal W. | Orchidée | Optimal W. |
| Superiority | 0.20 % | **24.40 %** | 0.80 % | **15.60 %** |
| Dominance | 1.60 % | **98.40 %** | 4.80 % | **95.20 %** |
| Converge | 43.60 % | **56.40 %** | 41.20 % | **59.80 %** |
| Diversity | **61.60 %** | 38.40 % | **58.40 %** | 41.60 % |

Table 18: Comparison of algorithms on the monophonic orchestration problems

In order to compare the performance of the algorithms, we also use the same evaluation measures Carpentier [71]. The four following measures are considered, where compared algorithms are named $\mathcal{A}$ and $\mathcal{B}$

Superiority We compute the dominated hypervolume of each algorithm which is not by the other, therefore $\mathcal{H}(\mathcal{A}, \mathcal{B})$ and $\mathcal{H}(\mathcal{B}, \mathcal{A})$.

Dominance The dominance measure is a combination of the *epsilon indicator* (which factor is required in order for $\mathcal{A}$ to dominate $\mathcal{B}$), the *coverage* (percentage of elements of $\mathcal{B}$ dominated by $\mathcal{A}$) and the *binary hypervolume* (hypervolume dominated by each algorithm).

Converge Defined as the mean of euclidean distances between each solution of the algorithm and the ideal point.

Diversity The spread of the distribution is computed over the optimization space.

We compare the original algorithm (*Orchidée*) Carpentier [71] and the *Optimal Warping* algorithm using these measures and report the results in Table 18 for monophonic problems.

As we can see, the *optimal warping* procedure strongly outperforms the standard multiobjective genetic (used in *Orchidée*) for the *superiority, dominance* and *converge* criteria. These results indicate that the optimal warping algorithms provide solutions that are superior to those provided by Orchidée. If we look more closely at the results, we can see that the *superiority* is almost never provided by Orchidée. These scores can be linked to the local optimization procedure, which enhance the (already efficient) solutions, by analyzing their performance in more details. However, if we look at the *diversity* criterion, we can see that Orchidée outperforms the optimal warping. This means that, even though the solutions provided by our algorithms are more efficient than the previous system, they are also less spread over the optimization space. This results seems to be a logical consequence of the local optimization procedure. Indeed, as we select and enhance only a portion of the Pareto front, the newly introduced solutions will concentrate around these areas of the optimization space. Therefore, this leads to sets that will cumulate in specific regions of the space. We provide in Table 19 the results of the performance assessment on polyphonic problems. We can see that the results are consistent for polyphonic mixtures. Therefore, the overall behavior of both algorithms is identical.

| | Polyphonic constrained | | Polyphonic unconstrained | |
|---|---|---|---|---|
| | Orchidée | Optimal W. | Orchidée | Optimal W. |
| Superiority | 1.20 % | **14.20 %** | 0.20 % | **19.20 %** |
| Dominance | 3.80 % | **96.20 %** | 2.40 % | **97.60 %** |
| Converge | 39.40 % | **63.60 %** | 41.00 % | **59.00 %** |
| Diversity | **61.0 %** | 39.00 % | **57.60 %** | 42.40 % |

Table 19: Comparison of algorithms on the polyphonic orchestration problems

### 12.4.4 *Modular structure*

The new system has been implemented around an *Object-Oriented Programming* paradigm. Therefore, the system is an extensible and modular structure which can allow modification, extension and introduction of tasks and search methods. Figure 70 summarizes the modular structure and components of the *ATO-MS* system. The *Session* object centralizes every information about a current orchestration problem and contains current instances of every sub-part of the problem. The *Production* object informs the system on the current means of sound generation that can be used (capacities and set of instruments that are allowed for a specific orchestration). The *Knowledge* object is used by the system to retrieve the symbolic and spectral features for individuals used in the search process (SQL database system). The *Search* module is the core of the orchestration system. These objects represent algorithms that are able to provide solutions. It can currently be set as a *SearchGenetic* (multi-objective genetic algorithm) or the *SearchOptimalWarping* presented in the previous sections. The *Target* defines the objectives to attain and therefore which values to approximate for each feature. As explained previously, the target can either be a *TargetSound* or a *TargetAbstract* that allow to directly input a set of features. The *Features* objects allow to define for each spectral feature its computation method, its prediction function and how to compute the distance between two elements. Finally, the *Population* defines the current set of proposals computed by an algorithm for a particular orchestration problem. Therefore, it contains a set of *Solution*s each of which is a mixture of several *Individuals*.

### 12.4.5 *Database*

The database in our orchestration problem is a representation of the extent of knowledge over the acoustic capacities of instruments. As our problem is fundamentally temporal, we include in this database a wide collection of temporal playing modes such as *glissando*, *crescendo* and so forth. We provide in Table 20, a comparison of the orchestral databases used by *Orchidee* and *ATO-MS*.

### 12.4.6 *Interface*

All the presented interactions and search algorithms have been implemented in a client/server architecture. The final interface for the ATO-MS orchestration software is depicted in Figure 71. As we can see, the interface provide a *server* communication window, which handles all the computation parts. Through the *target* panel, the users can input sound files or directly draw the temporal shapes of audio features to optimize.

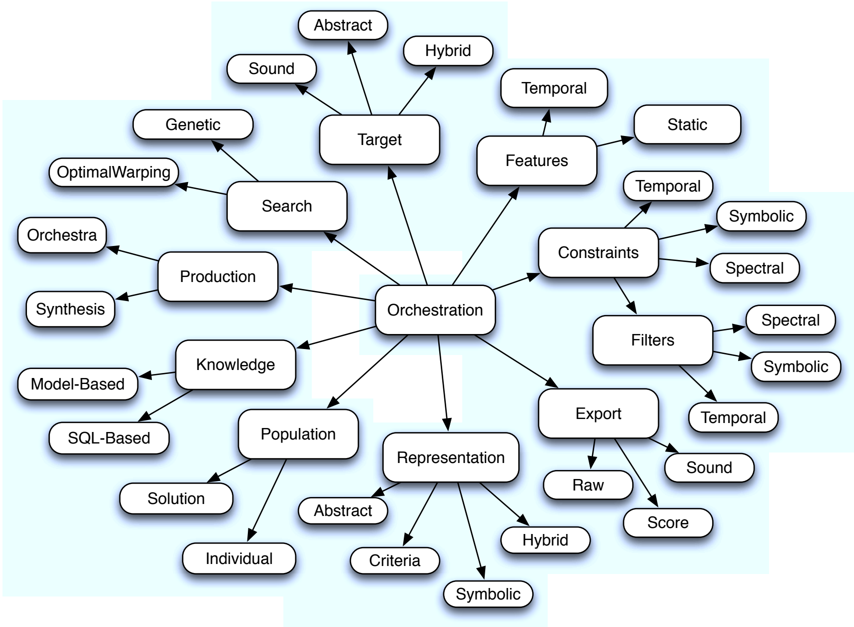



Figure 70: The current prototype for Abstract Temporal Orchestration with Modular Structure (ATO-MS) features an extensible architecture of modules to tackle the problem of Computer-Aided Orchestration.

Several sub-panels allow to control the *orchestra* to be used, the *database* interactions and other capabilities. When an orchestration search is launched, the panel on the right displays all the information on the found solutions. The optimization space allows to browse different sound mixtures based on their distances in different objectives. For each solution, a score displays the symbolic notation corresponding to the individual. Therefore, for each instrument it indicates the notes, dynamics, playing modes, onsets, durations and the corresponding sound samples that are used. The user can listen to previews of each solution and also have access to their predicted time series features.

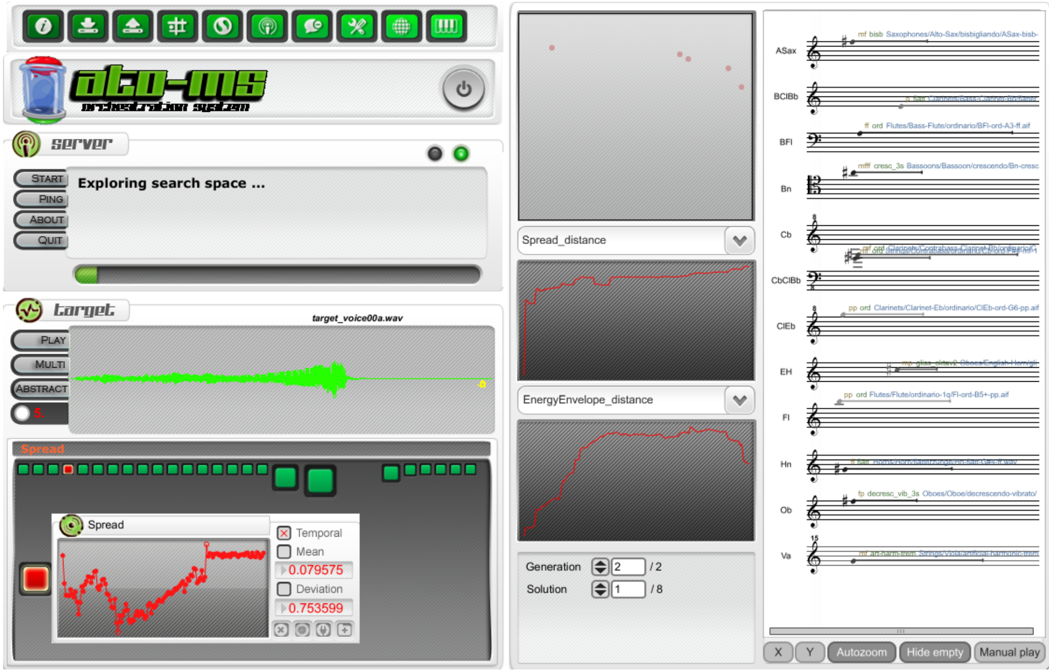



Figure 71: Interface of the *ATO-MS* orchestration software.

| | *Orchidée* | | *ATO-MS* | |
|---|---|---|---|---|
| Family | Modes | Samples | Modes | Samples |
| Bassoons | 8 | 458 | 18 | 1.359 |
| Clarinets | 18 | 814 | 37 | 3.959 |
| Flutes | 20 | 1.191 | 47 | 2.946 |
| Horns | 11 | 616 | 26 | 1.507 |
| Oboes | 11 | 695 | 33 | 2.950 |
| Saxophones | 2 | 161 | 22 | 517 |
| Contrabasses | 21 | 2.498 | 37 | 3.754 |
| Viola | 25 | 3.017 | 50 | 5.176 |
| Violins | 26 | 2.911 | 51 | 5.912 |
| Violoncellos | 25 | 2.930 | 49 | 4.549 |
| Trombones | 20 | 1.187 | 32 | 3.481 |
| Trumpets | 18 | 795 | 41 | 2.191 |
| Tubas | 9 | 891 | 21 | 2.064 |
| **Total** | **215** | **18.164** | **464** | **40.365** |

Table 20: Comparison of the orchestral databases used as a knowledge source for *Orchidée* (left) and the *ATO-MS* system (right).

# 13 OTHER ARTISTIC APPLICATIONS

We show in the following sections, the implementation of artistic applications of the MOTS framework. All the interactions and paradigms presented in this document have been implemented through various clients for different computer music softwares.

## 13.1 MOSEQ INTERFACE

We present in Figure 72 the interface for the MOSEQ paradigm, implemented as a Max/Msp client. The user can select a set of features to optimize, displayed under drawing boxes. The users can draw the temporal evolution of the corresponding features so that it closely match their idea. When subjects are satisfied with their drawings, they can perform a query. The MOTS algorithm will provide a set of solutions spread over the optimization space. Users can listen to the results, see the temporal evolution of their features and try to find their sounds. When the user puts the mouse over a solution, its set of features is automatically displayed. The user can listen to the sounds by clicking on them.

## 13.2 QVI INTERFACE

The interface for the QVI system is presented in Figure 73. The users can here perform a vocal imitation of their search. The complete set of sound features is displayed in real-time while the subject is recording. After recording their imitations, the users can modify the temporal shapes by direct input. When the users are satisfied with their input, they can perform a query. The MOTS algorithm will display the corresponding set of tradeoffs solutions. When the user puts the mouse over a solution, its set of features is displayed. The user can then listen to the sounds by clicking on them.

## 13.3 OPENMUSIC CLIENT

We developed an OpenMusic client that embeds the querying possibilities presented previously. This client is presented in Figure 74. Hence, it allows to perform temporal queries as well as MOTS, MOSEQ and QVI queries. This client also show the possibility of adding symbolic constraints over temporal queries in order to restrict the search space.

## 13.4 IPAD INTERFACE

All the aforementioned technical characteristics are also implemented as a multitouch interface for the iPad. This interface embed all interaction schemes and capabilities in an OpenGL / Objective-C framework. Therefore, it provides the MOTS, MOSEQ and QVI paradigms for interacting with a temporal audio database. It has been developed as a client to a local computer server (which allows an extensive storage size) using OSC communication but can also be self-contained (even if the database is inherently limited). The main screen of the interface is showed in Figure 75.

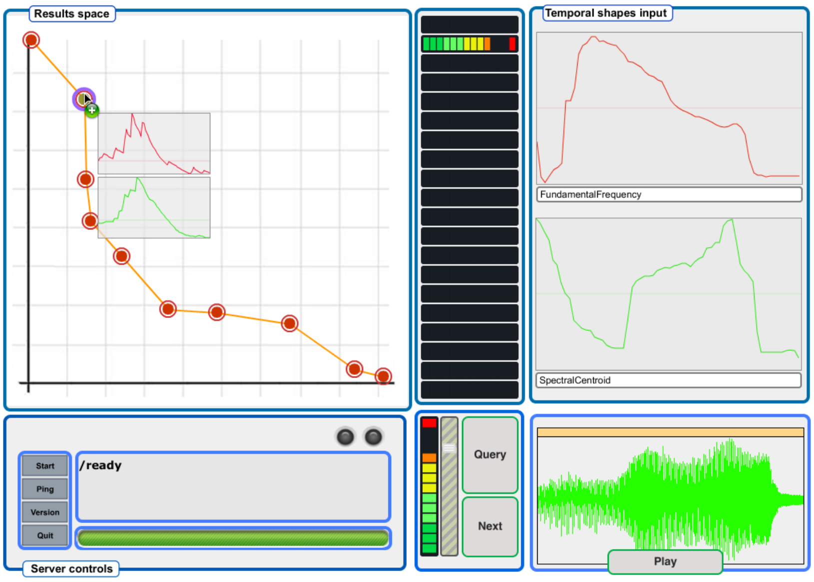



Figure 72: Interface for the MOSEQ system. The two current features are displayed under drawing boxes. The users can draw the temporal evolution of the corresponding features so that it closely match their idea. When users are satisfied with their drawings, they can perform a query. The MOTS algorithm will provide a set of solutions spread over the optimization space. Users can listen to the results, see the temporal evolution of their features and try to find their sounds.

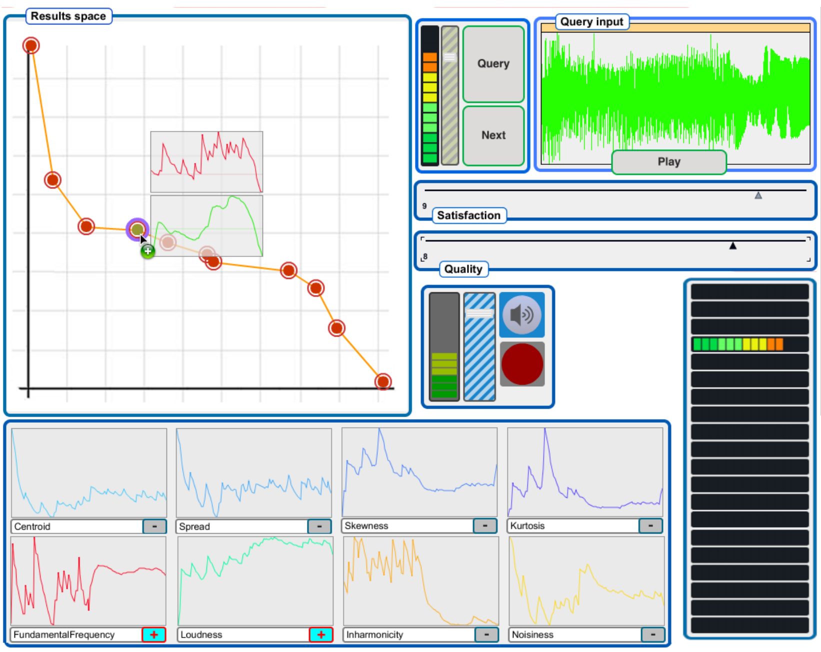



Figure 73: Interface for the QVI system. The users can then perform a vocal imitation of the target. The complete set of sound features are displayed in real-time while the user is recording. After recording their imitations, users can modify the temporal shapes by direct input. When users are satisfied with their input, they can perform a query. The MOTS algortihm will display the corresponding set of tradeoffs solutions.

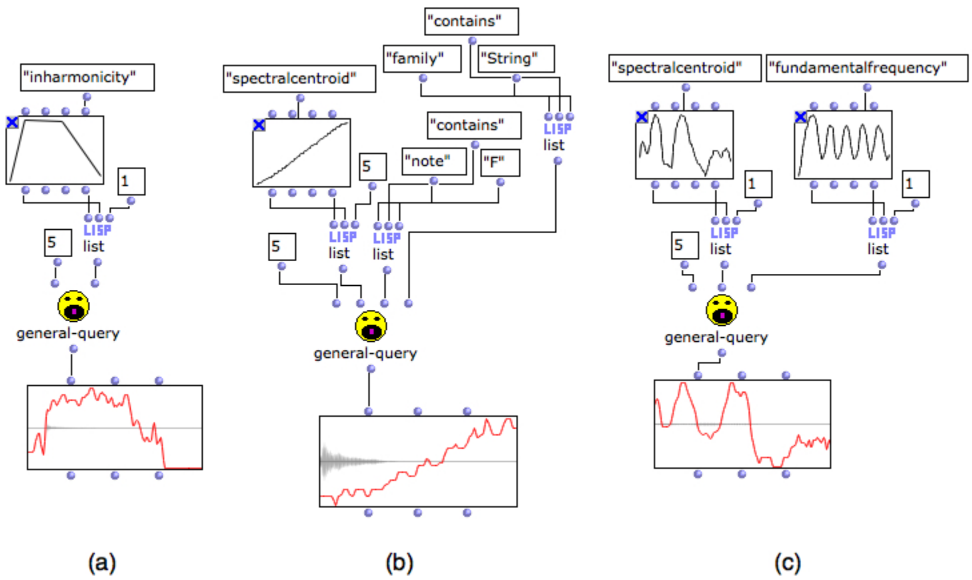



Figure 74: Client for the querying system in OpenMusic. This client also show the possibility of adding symbolic constraints on top of temporal queries in order to restrict the search space.

## 13.5 Spectral maquettes

We present here a preliminary system that could embed several composition paradigms. Its goal is to allow a link between the timbre and musical writing by imposing temporal relations on audio features. It has also been designed to embed symbolic and signal units in a common framework. The system relies on the time series database and audio querying paradigms that act as an organized lexicon. This system called *spectral maquettes* for OpenMusic (OM) allows temporal and structural interactions between musical units of various nature. It is the first step towards a wider composition framework that could navigate between symbolic and spectral data.

### 13.5.1 *Motivation*

During last century, instrumental music has undergone a turnaround towards increasingly inharmonic, noisy and strongly time-variating sounds. The advent of technology has pushed back the frontier between noise and music, as evidenced by the approach of Pierre Schaeffer [? ]. We can deplore the lack of compositional systems that could catch up to these new practices. Nowadays composition systems usually follow the traditional harmonic paradigm and are inscribed in a punctual time of writing. However, a contemporary issue in computer music research is the *signal / symbolic* interaction. On the one hand, the analysis and synthesis of digital sounds allowed the production of sounds previously unheard. On the other hand, the algorithmic approach focused on the symbolic structures of musical notation. The orchestration is located precisely at the crossroads of these two lines of research. If it claims to create timbres, it is through a writing process. Therefore, it is a meeting point between the symbolic and spectral domains. So it can be grasped by computer music only through a convergence of the signal and algorithmic approaches.

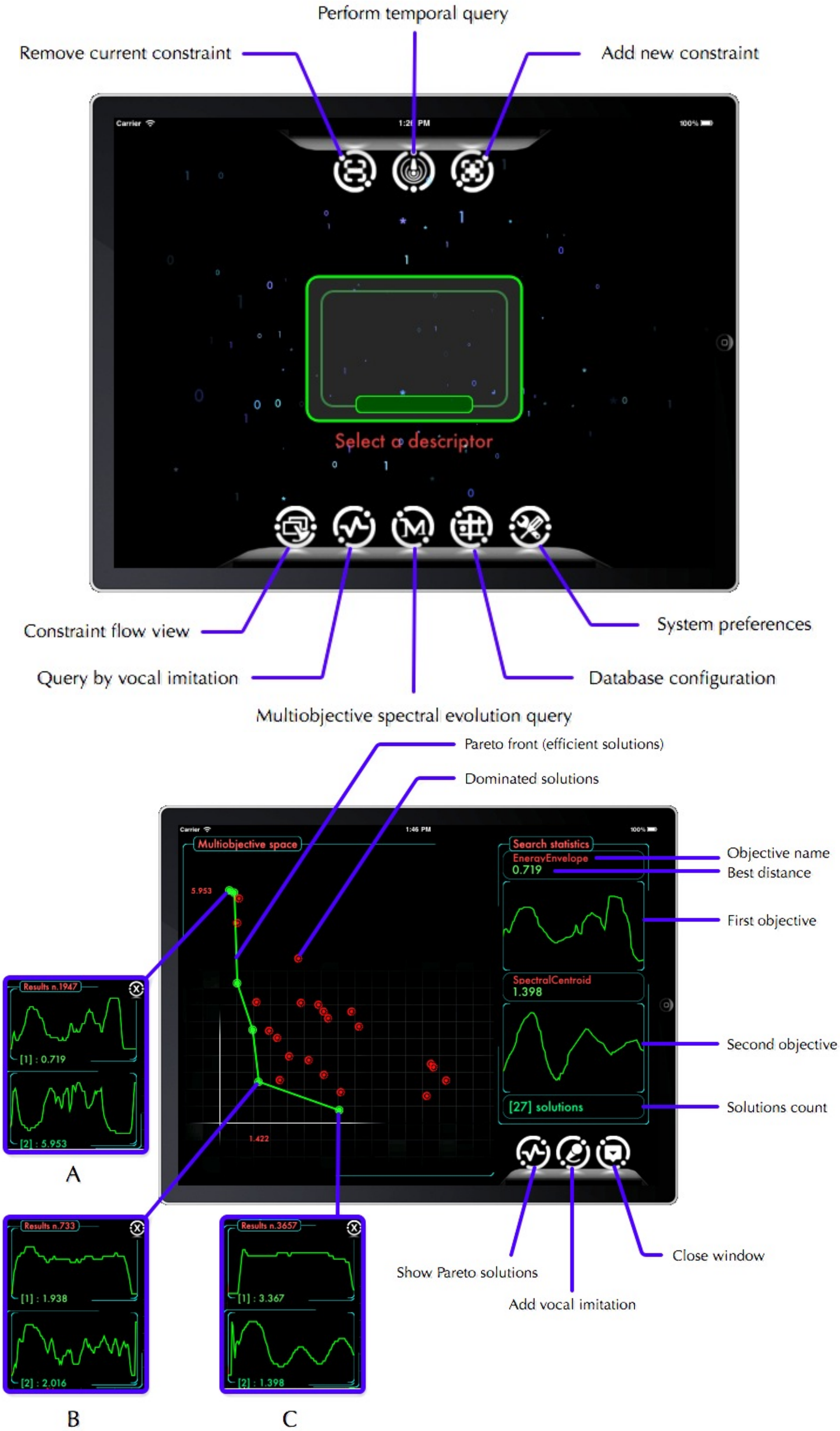



Figure 75: Main panels for the iPad audio database interface.

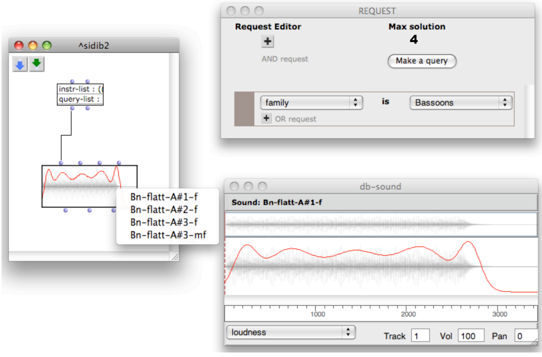



Figure 76: Make an instance of the `db-sound` class by a SQL query. The new `db-sound` instance contains four sounds. A special editor displays the current one and different descriptors of it.

We propose a composition system that could effectively deal with the stylistic aspects focused on timbre. We use the knowledge obtained with the MOTS framework through the previously presented database. The audio querying paradigms allow to perform queries based on the temporal evolution of descriptors rather than simply on static criteria. This architecture thus establishes an organized lexicon for our system. In order to compose with sounds at the level of the musical discourse, we use OpenMusic's maquette system as a basis. A maquette can be seen as a "meta score" embracing in a single document musical notation and visual programs.

### 13.5.2 *Implementation in OpenMusic*

First, we defined a new class of sounds belonging to the database (`db-sound`). User can make instances of this class by sending SQL queries to the spectral database (see Figure 76). More than a sound, a `db-sound` is a collection of sounds whose cardinality is specified by the user. As common sound files, `db-sound` instances are first class citizen (i.e. they can be included in visual programs, scores or other OpenMusic editors). By using `db-sound` instances along with the general visual programming tools of the OpenMusic environment, composers are provided with means to develop complete processes. In this context, the functional program structures let them manage the complexity by maintaining hierarchical control from musically relevant abstractions down to synthesis processes.

In particular we have implemented a new class of maquettes called `Spectral Maquettes`. The maquette is an extension of the notion of visual program with additional spatial and temporal dimensions. This allows to put the elements of the composition framework (data structures and processes) in close relation to these two dimensions. In a maquette editor, the boxes (called temporal boxes) represent functional units (programs) producing musical outputs. The position and graphical properties of these boxes are associated with a temporal and structural sense; particularly, the horizontal axis of the editor represents time, so that the position and horizontal extension can be related to offsets and durations. The temporal boxes can also be linked by functional connections so that the whole maquette may finally be considered as a program, comprising functional and temporal semantics. Temporal relations and constraints can therefore be set between the boxes by setting the temporal parameters in their corresponding patches. Hence, the calculus can determine the time structure.

In addition, the possibility of embedding maquettes in a maquette allows for the construction of hierarchical temporal structures.

The spectral maquette is an extension of the classical one aimed at fomenting the relation of musical material and processes with micro-time structures. Indeed, sound mixtures require a particular attention to the organization, sequencing, and articulations of a global musical form. We show in Figure 3 an example of spectral maquette. When a `db-sound` is put into the maquette the subjacent box shows the wave form of the sound augmented with several information like clue points (e.g. attack, sustain and release) or the evolution of some descriptor (e.g. loudness, noiseness, spectral centroid). Boxes into a spectral maquette can be related between them in different ways. In Figure 3 (a) we can see how the start time of a sound is linked to the first note of a melody; moving one of these boxes force to move the other one in order to respect this constraint. In part (b) three sounds are superposed, we can see in this example several types of temporal relations: the lowest one is attached to a flag in the temporal ruler, this sound will start always at this time; the middle box, in turn, finishes the lowest one (other Allen's relation can be easily imposed between boxes); finally relations between the highest box and the middle one are defined by means of internal points of sounds (these points can be defined by hand or coming from queries on the database). In part (d) the `db-sound` is taken as an input of a program which builds a sequence of notes whose melodic profile follows the evolution of the loudness of the sound. Finally in part (e) a viewer allows to select a particular descriptor and calculate the result of the superposition of all `db-sounds` in the maquette. In a first time this viewer allows to see the global evolution of the superposition, but it is not difficult to imagine that by changing the curve the user can launch a query that search for new boxes approaching this new curve. Hence, the spectral maquettes enable to structure the musical information at different levels :

- the static level of the form, i.e. the organisation of boxes in time.
- the dynamic and paradigmatic level of the form (i.e. constraints between the temporal boxes).
- the syntactical level, i.e. the calculus building the musical discourse inside the temporal boxes.
- the material level, i.e. `db-sounds` taking part in the maquette

These four levels of information are obviously interconnected. The most important advantage in the spectral maquette concept is to offer a visualisation of this interaction and at the same time, interactive control of it. This produces a source of exploration and experimentation :

- Recombination at the form level. Temporal boxes are moved and stretched in time without changing the other three levels.
- Modification of functional relations. We do not change the position of the blocks but their causal relation.
- Syntax modification. Algorithms which build the material can be changed according to the compositional goals.
- Change of the material by launching new queries to the database.

When combining these procedures, sophisticated musical experiments can be performed by the users.

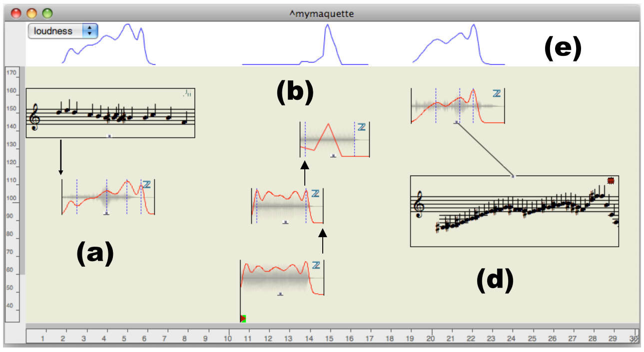



Figure 77: Concrete example of the usage of spectral maquettes. Several functional, macro and micro-temporal relations are defined between boxes.

# Part VI

# CONCLUSIONS

# 14 FUTURE WORK

## 14.1 THE MOTS PARADIGM

### 14.1.1 *Wider applications*

We believe that the framework of MOTS matching could lead to powerful applications. This algorithmic problem has, to the best of our knowledge, never been addressed. Hence, we are expecting to develop its application range. We already tried with the subsequent HV-MOTS *classification* to provide a maximal range of application fields. However, the idea of *matching* the time series in a multiobjective manner can also provide an interesting variety of applications. In this section, we envision some potential applications for future work.

*Sound computing*

As the structural choices underlying the MOTS framework were drawn from observations of our auditory perception, it would appear logical to extend its scope of audio applications. In this study, we already applied the MOTS system to the analysis of sound samples and shown that it was fit for both matching and classification. However, an interesting application would be to extend this scope to musical song matching. For example, a dual high-level analysis can lead to obtain a disjoint set of rythmic and melodic time series for a same piece. Therefore, the multiobjective flexibility could also allow to separately match rythmical patterns and melodic series on a larger time scale.

*Medical diagnosis*

The information gathered for medical analysis is usually provided by sensors that collect repeated measurements at fixed intervals of time. Therefore, they produce time series observations such as electrocardiogram (ECG), electroencephalogram (EEG), blood pressure and so forth. These signals are mostly used to monitor the state of health of patients. Hence, one of the major concern of medical informatics has been to help diagnosis and prognosis. Because of the temporal nature of medical signals (and the wealth of source where measurements could be extracted), an interesting line of work would be to study the application of the MOTS framework in these fields.

*Genetic analysis*

In bioinformatics, a lot of research has been devoted to the analysis and matching of DNA sequences. These sequences can be considered as time series, either in their raw format or by transforming these into random walk-type series by putting different weights on nucleic bases. Furthermore, there also exists a flourishing literature in multiobjective optimization for computational biology. Therefore, the field of genetic analysis appears as a promising topic of study for our algorithmic framework.

### 14.1.2 *Hybrid analysis*

Along this document, we studied the application of flexible notions of multidimensional similarity to sets of time series. However, an interesting line of future work could be found in the expansion of these concepts to even more complex problematics.

*Combining views*

First, an interesting line of work lies in experimenting some variations on the concepts of multiobjective similarity. For example, the normalization of time series data allows to separate its shape from its first statistical moments. Therefore, in the context of audio applications, it appears interesting to perform queries that jointly minimize the mean, standard deviation and temporal shape of the same feature.

*Multiobjective subsequence matching*

The main idea of the MOTS framework is analyze the similarity on sets of *distinct* time series. However, a very interesting study would be to apply the same concept to studying the similarity of *subsequences* of the *same* series. Therefore, this would allow to obtain several objectives that rely to the *same time series*. Hence, this study would extend the MOTS framework to work *inside* univariate time series. This relates to the stimulating question of time scales continuum, that we will detail in Section 14.8.

### 14.1.3 *Interaction and representation*

Multiobjective analysis have been applied to a wide range of problems in the past years. However, a traditional problem lies in the representation of the solutions and the possibilities of interaction offered by different systems. Indeed, multiobjective analysis can span a very high-dimensional space. Therefore, the representation of the Pareto solutions is bound to provide only a dithered view of the optimization space. Therefore, an interesting line of future work would be to study the potential solutions for representation and interaction with the MOTS framework. A good premise to this study would be to study dimensionality reduction techniques that could be applied and how these could influence the design of interaction schemes.

## 14.2 MOSEQ / QVI

### 14.2.1 *Applications in audio workflow*

The two audio querying paradigms introduced in Section 7.5 looks promising given the results of their user validations. However, their use in a seldom context do not appear relevant to nowadays audio workflows. Indeed, the current music production architectures are converging towards fully integrated framework. Sometimes, these even provide video edition features or other unrelated supplements. An interesting direction of research would be to study how the MOSEQ and QVI paradigms could be integrated in a music production workflow. Furthermore, their integration in such frameworks would require a form of interaction with other musical modules. Therefore, it would be interesting to study how the inputs and outputs of these paradigms could be generalized towards natural integration.

### 14.2.2 *Relevance feedback*

As we discussed in Section 7.5, the multiobjective space embed a notion of *preference* towards different solutions. Indeed, each solution in the Pareto front is correlated with an underlying set of weights. Selecting one of these solutions rather than others implies a clear preference over different optimization dimensions. Therefore, an interesting direction of research would be to provide a weighting mechanism for each dimension in order to find relative preferences. Then, this approach could be extended to compute persistent "global" weights from one search to another. Finally, this could lead to automate this relevance mechanism. In this line of thought, an interesting user study would be to analyze if subsequent searches regularly goes towards the same optimization directions. This idea could ultimately bring an auto-modifying framework that could intelligently adapt to different users.

## 14.3 HV-MOTS CLASSIFICATION

### 14.3.1 *Audio applications*

The HV-MOTS classifier isn inspired by concepts from the auditory perception. Therefore, it would seem logical to apply it further in different fields of audio applications. Several topics of MIR are fundamentally directed towards classification problems and could benefit from this new classification scheme. For example, it would be interesting to apply the HV-MOTS classifier to wider time scales and elements of study. Furthermore, several classification problems are still open and would be interesting to study with the flexibitlity of this framework. These topics comprise *musical genre* classification, *musical mood* inference and *music composer* classification.

### 14.3.2 *Scope of application*

As we have seen in Chapter 9, because of the ubiquitous nature of time series, the HV-MOTS classification paradigm can be applied to a wide range of problems. We tried to retrieve a maximal diversity of datasets on which to analyze the accuracy of this framework. However, as put forward by Demsar [102] even the largest scale studies may not be able to project results on datasets that were not part of the original study. Therefore, an interesting direction of future work would be to analyze the statistical superiority of the HV-MOTS classifier on an even larger scale. We believe that the flexibility of this classification framework could prove itself to be statistically superior on a wide range of topics. In fact, multiobjective optimization techniques and time series analysis have a long history of successful but separate applications. Hence, an interesting line of future work would be to analyze the introduction of time series in multiobjective analysis framework. Convertly, the multiobjective flexibility could enhance traditionnal results in time series classification. Therefore, the HV-MOTS classifier could still be applied to a whole range of scientific fields comprising *economics, chemical engineering, process design, scheduling, bioinformatics, computational biology, climate analysis* and *medical diagnosis*

### 14.3.3 *Multiobjective subsequence classification*

As we argued along our study, the main property of the HV-MOTS classifier is that it is constructed to provide multidimensional similarity assessments. Therefore, it is theoretically meaningless to apply it to univariate time series classification. Indeed, in this case the hypervolume selection provides a classification decision equivalent to the 1-NN classifier. However, an interesting line of future work could arise from the study of *multiobjective subsequence classification.* A first approach to this concept would be to try to extract several objectives from a set of univariate time series related to the same feature. Therefore, it would allow to study the accuracy of the HV-MOTS paradigm in univariate time series classification. However, this requires to study how to extract those subsequences, choose the corresponding cutting points and, of course, its computational complexity.

## 14.4 HEART SOUNDS BIOMETRY

As we have seen in Chapter 10, the S-Features and HV-MOTS approach allow to construct an accurate identification system based on heart sounds. Even if the error rates already outstands the state-of-art proposals, a wide amount of work can still be performed to enhance these results.

### 14.4.1 *Segmentation procedure*

As we have seen, the segmentation procedure is one of the most critical module when analyzing the final results. It seems that the tuning this procedure may dramatically impact the error rates. Therefore, one of the most important enhancement for this system would be to provide a segmentation procedure directed towards the specificities of heart sounds. Indeed, our segmentation procedure has been selected inside a set of available methods for sound segmentation. However, more specific techniques could be developped, with a special interest towards the use of the Stockwell transform as the basis for segmentation.

### 14.4.2 *Features computation*

We developped a specific set of features based on the Stockwell transform (cf. Section B.2) for heart sounds biometry. It seems that the S-Transform is a valid choice as it allows to outperform classic tools of features analysis for this specific problem. However, this set still seems somehow limited compared to the wealth of research that has been performed in the field of audio analysis. Therefore, an interesting line of work would be to expand the set of S-Features for heart sounds biometry. Nevertheless, another interesting work would be to analyze the applicability of this feature set to other problematics. Of course, a natural choice would be to try to apply this set to audio problems. However, because of the low-frequency oriented nature of this feature set, it might also be relevant to other problems.

### 14.4.3 *Factors of influence*

We tried to list in Section 10.5 the factors that could influence the performances of a heart sounds biometric system. Therefore, it would be of utmost interest to study the true impact of these factors on the performance of the system. The most interesting parameters to study would be the *physiological factors* and *heart diseases.* However, the success of such analyses requires the collection of more exhaustive datasets. For example, studying the *physiological factors* in a population of subjects imply that we should record the *same* subjects in varying physiological conditions (rest, light and hard physical exercise, sleep). Therefore, a first mandatory step would be to perform an exhaustive data collection campaign. This would allow to obtain databases that could compete with nowadays knowledge on other biometrics.

## 14.5 On heart diseases detection

After analyzing the results of the heart sounds identification system, a natural idea may come to mind. If the system is so efficient in identifying persons through the sound their heart produce, it should easily detect heart diseases based on the same information. Indeed, as this approach is able to discriminate extremely small variations of the heart sound signals, it should be even easier to discriminate between extremely varying signals such as heart diseases. However, as for the previous section, such a study would require to collect a very wide dataset of heart diseases to support a valid statistical conclusion. Ideally, the dataset should even contain some subjects that are recorded *before* and *after* developing a heart disease. Obviously, it appears to be a very hard case to find but also implies numerous ethical issues.

## 14.6 Orchestration

We proposed in Section 12.4 a new system for *temporal* and *abstract* computer-aided orchestration. We have shown that it provides a powerful enhancement over existing systems. However, as we discussed in Chapter 12, musical orchestration is an extremely complex topic which encompass several open issues and problematics. We outline here some potential trends and avenues for the next years of research in musical orchestration.

### 14.6.1 *Signal and symbolism*

Musical orchestration offers a unique framework to study the interactions between the signal and symbolic research streams. Linking the symbolic world of writing with the realm of signal could lead to potentially powerful applications. We will discuss a hypothetic system implementing these ideas as an interesting line of future work in Section 14.7.1. We focus here on performing an analysis of such ideas through the existing body of musical works.

#### *Relations between scales*

In order to find a relation between the signal and symbolic layers, it would be interesting to analyze existing musical pieces through an automatic knowledge extraction procedure. Therefore, this analysis would focus on finding musical configurations where links between spectral patterns and symbolic writing are denoting a logical joint evolution. This study would require a coordinated analysis between the score, instruments information and resulting audio features. This joint analysis could identify the relationships between these two worlds. In order to link the acoustic time series and symbolic objects, an interesting lead would be to use artificial intelligence technique between a set of signal features and symbolic information.

*Extending to generic sound mixture*

Even if classical orchestration is still an open problem, an interesting avenue of research would be to extend it to generic sound mixtures and electronics. This requires an environment design which could take into account the specificities of electronic parts. This goal is already being assessed by composers through mixed orchestration situations. In these contexts, we need to find configurations in which these two acoustic modes can blend together, without giving an impression of two dichotomic worlds. This problematic can be approached by subverting the problem of orchestration to the problem of timbre

### 14.6.2 *Intelligent music notation*

Current music notation systems does not support the *contextual rules* of music writing. Hence, they provide no further analysis of playing techniques and acoustic properties of individual instruments. While sophisticated word processors feature aiding tools to verify *grammatically correct* texts, no current music notation systems offer such "musical grammar" aid. Particularly in the context of orchestration, musical writing can be a painstaking task. Indeed, careful considerations of fingering, breaks in register, masking effects and so forth should be taken into consideration. Therefore, an interesting line of work would be to develop music notation systems, capable of recognizing incorrect writing and recommending alternative corrections. Therefore, a large amount of knowledge should be gathered and represented efficiently on both the symbolic and spectral aspects of instrumental capacities. This knowledge gathering procedure would require an amount of manual retrieving to ascertain the validity of information. It should also provide acoustical properties of the instrument and how these can affect questions of orchestral effects.

*Automatic knowledge gathering*

Based on the previous goals, a line of future work would be to automate the gathering of instrumental knowledge for orchestration. The idea would be to construct this knowledge incrementally from an automatic analysis on existing scores. Starting from a "*nothing is possible*" configuration, each apparition of a particular symbolic sequence adds it to the knowledge, and recurrence makes the sequence easier to reproduce (on a probabilistic scale with a confidence degree). Based on the symbolic scores, each sequence, chord and fingering can be added in the knowledge database and considered feasible. The more instance we found in analysis, the easier it should be to perform the related sequence. Then a cross-analysis with corresponding signal recordings and audio features could allow to obtain deeper and more interesting relationships between symbolic and signal descriptors. This approach could lead to a powerful knowledge database on signal/symbolism relations.

### 14.6.3 *Emergence phenomenon*

Orchestration embeds some of the most misunderstood acoustic phenomena. First, the blending of orchestral instruments can almost make it impossible to find back these original constituents (a phenomenon termed as *orchestral fusion* in the literature). Therefore, the orchestral timbre contains complex non-linear interactions. Several phenomenon have been observed in music listening and are termed as *emergence phenomenon*, but are yet to be fully explained. First, the well-known *masking effect* in which some elements disappear from our perception, the *chorus* and *unison* effects (in which a group of the same instruments play the same notes) and the *rugosity* and *coupling* effects. It seems crucial to understand these phenomena in order to integrate them in the orchestral reasoning. Therefore, we should study some groupings of instruments which generates some unexpected elements inside the spectrum. A first solution would be to obtain recordings of several emergence phenomena and try to find the relation between their spectrum.

## 14.7 closing the gap between all worlds

As we have talked extensively in Chapter 1, orchestration lies at the crossroads of signal and symbolism but also thrives between micro-temporal evolution and macro-temporal articulations. We tried to sketch a first approach that could help in translating the intent of a composer in the process of orchestration through the *spectral maquettes*.

Hence, an interesting line of future work would be to allow a more intuitive and therefore easier control of sound mixtures. This system should provide an nteraction between the spectrum (spectro-) of sounds and their evolution in time (-morphology). However, although the spectral content and temporal evolution are undoubtedly linked, we need to separate concepts to describe them. Based on these remarks, we envison a first system that could try to bridge the gab between those worlds through the concept of constraints inference.

### 14.7.1 *Constraint inference system*

The main idea behind this system would be to extract the implicit relationships that are created by a composer while writing music, as exemplified in Figure 78. For example, a chord will surely be related to its context, ie. previous notes. However, we can also find some spectral and temporal relations inside that chord. Ultimately, the goal would be to let the users compose freely and then obtain a constraint network that could be modified algorithmically. That way, we can infer some high-level relations as well as micro-level properties. By processing a simultaneous analysis of the symbolic score and the corresponding audio features evolution, we could obtain a graph of constrained relationships, explaining the link on both types of viewpoints. The key here is in combining time series inference and knowledge mining approaches. For instance, by using a spectral database, it is possible to perform a musical writing and at the same time predict the audio features of each melodic line. Therefore, the links between the harmonic rules and signal features could be explicited through a constraint network (as shown in Figure 78) Then based on the infered network, it would be possible to modify and then solve alternate networks. The main problem for the inference system would be to filter the knowledge induced from an automatic constraint inference. Indeed, some of the infered constraints could be irrelevant. For example, note X is played 143,35 seconds after the beginning of the piece seems like a rather dull information. Oppositely, the temporal relations between this note and others are crucial to the final musical work.

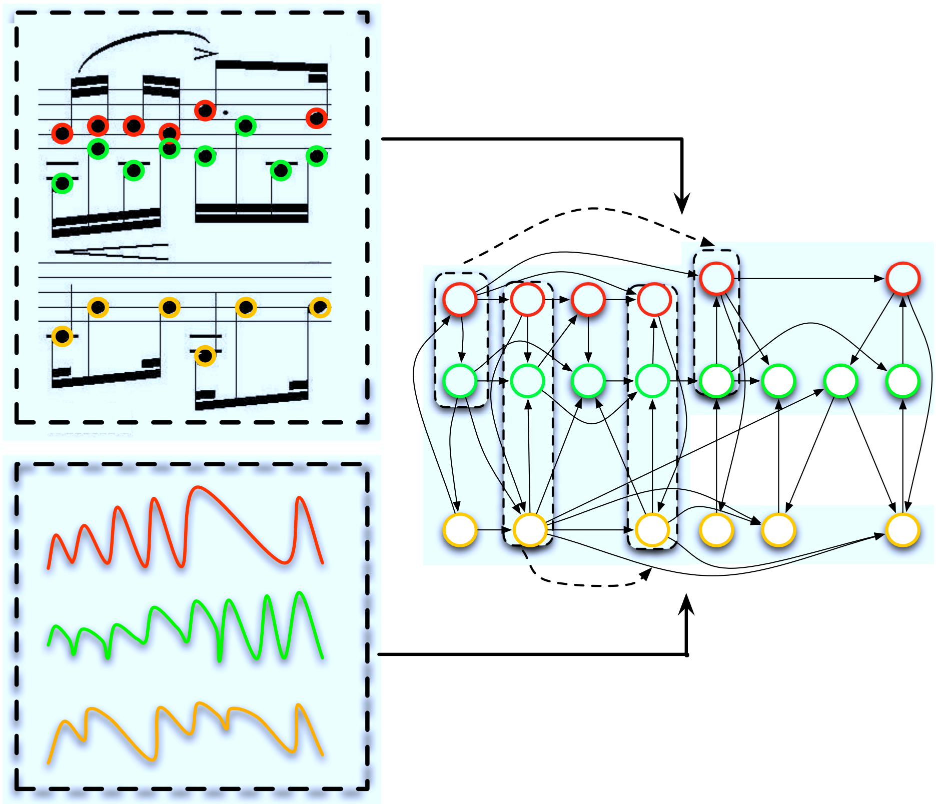


Figure 78: A complete system of constraint inference that allow to bridge the gap between the symbolic realm of musical writing and the signal world of timbre. A simultaneous analysis of the symbolic score and corresponding audio features evolution could provide a graph of constrained relationships, explaining the link on both types of viewpoints.

### 14.7.2 *Several views on constraints*

This constraint system could be articulated over several aspects of an orchestration system. Therefore, it should constitute a multi-level hierarchical system. We try to list here every beneficial module that could lead to an unifying framework of constraint handling.

- *Instrumentation and static constraints* - Symbolic constraints between instruments, notes, playing modes or other symbolic parameters are a first extensive topic of study. Instrumentation constraints on fingering, articulations and writing could exhibit links between successive orchestrations on the symbolic level. (eg. *continuity*, *diversity* and so forth). In fact, a tremendeous amount of work on instrumentation is yet to be done (we detailed some research avenues in Section 14.6.2). This approach requires an extensive knowledge database with fingering, articulations speed and every instrument technical capacities.

- *Temporal constraints and macro-articulations* - We can draw several advantages from linking a purely compositional writing framework with a more generic spectral orchestration system (as ATO-MS presented in Section 12.4). This brings back the link between different temporal scales from the micro-temporal evolution of signal properties to the macro-temporal articulations of musical discourse (we will try to develop this avenue of research through the idea of *time scales continuum* in Section 14.8). In terms of constraints, this would imply to establish a "meta-level" of constraints to provide a continuous link between several temporal scales. Regarding macro-articulations a straightforward example of application would be to find *orchestral paths* between two orchestrations (examples of constraints include "*minimize the number of instruments change*" or "*maximize the variation on a set of descriptors*")

- *Formal orchestration constraints language* - An interesting line of work would be to define an operational language for orchestration by formalizing its basic operations. This would provide a formal language for orchestration by starting with simple operations and then complexifying these by iterative refinement. Therefore, this language should be defined from its constituting elements (*musical atoms),* combined through simple operations (eg. amplification, transposition, overlay) and towards a constraint checking and solving procedure. The first step is to characterize the language primitives that could range from sinusoidal components to complete melodic lines (polymorphism of objects). The set of operations should also encompass this range between signal operations (amplification, transposition and stretching) to symbolic processing (concatenation, overlay, alignment and assembly). Finally, the operators should also embed the different time scales through a notion of *operator granularity.*

## 14.8 ON A TIME SCALES CONTINUUM

We have seen along this study, that we can work with widely varying time scales, each of which provide a different information focus. We try here to foresee the basis for a more generic problem. We hypothesize a more *elastic notion* of time based on a *continuum of time scales*, which could handle the notions of *temporal granularity*. This question rise from the simple fact that we can not define a clear boundary between the notions of *micro* and *macro-time*. For instance, a continuous glissando of several seconds is considered as a micro-temporal evolution. However, in the same time frame, we can construct complete melodic lines for different instruments. These are obviously considered as a macro-temporal articulation of musical discourse. Therefore, in musical problematics at least, there seems to be a discrepancy of the classical notion of time scales. Based on the idea of scales continuum, it would be interesting to extend our reasoning to a dense set of temporal domains. This would allow to perform a similar reasoning at every time scale. Hence, processing each temporal granularity in the same manner would provide a wider homogeneity. We discuss this problematic through the prism of orchestration but it could easily be transposed to any type of temporal process.

### 14.8.1 *From micro to macro*

The connection between micro and macro temporalities would allow to incorporate timbral structure in the compositional act. Hence, we must find ways to reassamble the temporal reasoning from the smallest levels all the way up to macro-temporal level. A first step would be to study the states of interaction between different temporalities and how the micro-time can influence the macro-level. This leads to a *"problem of uncertainty"* in which the relationships between local and global representations of time-varying spectrum are yet to be connected. A solution could be to use of a bottom-up pattern recognition process based on small collections of primitives, that could result in clustering of features into patterns. It is also interesting to consider the creative part of parsing in its connective ability (and not merely reducing) that can encompass several layers of description. This approach would seek to organize these layers into an integrated musical structure with multiple facets. It is unlikely that this task is looking for a 'one-to-one' correspondence between the layers. It should rather recognize that the musical context emerges from nonlinear interactions. Dimensional reduction can give a good angle to attack this problem. It would allow to analyze the evolution of musical structure and how this movement is correlated with its spectral rendering. There may be no linear continuity between the different time scales. However, the solution can not consist of the outright closure of their borders, but should consider all the possibilities of interaction and hence articulation between these different scales. This question of time granularity could connect the treatement of events from different time scales in the same encompassing framework.

### 14.8.2 *Macro-temporal articulations*

The second step of this problematic lies in defining the evolution of sound mixtures over large period of time. As we discussed earlier, the temporal constraints can define paths between several orchestral states. An interesting possibility would be to provide symbolic paths and arbitrarily choose the temporal evolution of a spectral dimension to be optimized. Therefore this study should try to find the link that can unite the spectral flow of two orchestrations in a temporal manner. Therefore, it would be interesting to study long-term modes for "articulated" targets whose timbre is changing continuously.

# 15 CONCLUSIONS

We have shown along this study that a musical problematic can give birth to powerful analysis schemes, far beyond their own study. By questionning the nature of artistic reasoning and at the same time drawing inspiration from our musical perception and gaining insights from these mechanisms to drive our choices of algorithms, we have been able to create novel and powerful approaches for generic querying and classification.

First, we introduced the problem of *MultiObjective Time Series* (MOTS) matching and its formalization. We discussed the core differences between this novel framework and its multivariate counterpart, and have shown how it could lead to more flexible sets of retrieval solutions. We introduceds two efficient algorithms to solve this problem by relying on the concept of *approximate hyperplane.* We have analyzed their relative merits on synthetic and real datasets, which exhibited that the *hyperplane* algorithm was able to solve the MOTS problem in sub-linear time with respect to both database and objective cardinalities. However, further analysis of the results show that the variance of querying time are more important with this method as its complexity depends on the distribution of data. Based on the MOTS framework, we were able to formalize innovative audio querying by introducing two new querying problematics in the field of audio samples retrieval. The *MultiObjective Spectral Evolution Query* (MOSEQ) in which users can directly draw the temporal evolution of audio features and *Query by Vocal Imitation* (QVI) which allow users to perform a vocal imitation of the sound they are seeking. Both paradigms are a direct application of the flexibility introduced by the MOTS matching problem. We then provided a discussion on user studies for the temporal evolution of audio features. These studies are aimed to validate the hypotheses on which the MOTS framework, which exhibited the concept of *directions of listening*.

Based on these results, we extended our scope of study and showed how to apply these notions of flexible similarity evaluation to classification problems by introducing the HyperVolume-MOTS (HV-MOTS) classification scheme. We showed that even within the multiobjective framework that avoids merging distances into a single measure, we can still rank classes by relying on the hypervolume dominated by each. We discussed the relationships between this novel classification framework and other distance-based classifiers but also to more generic classification schemes and further provide a discussion on its main advantages and drawbacks. We provided a large scale study of the performances of this classification technique on a wide range of datasets that covers several scientific fields. We showed the statistical superiority of the HV-MOTS classifier over well-established classification schemes and over state-of-art results on the same datasets. Based on the HV-MOTS classifier, we showed how to construct a biometric identification system for heart beat sounds by developing a specific set of features based on the Stockwell transform, called *S-Features*. We showed that using heart sounds as a biometric feature provide a reliable identification. We provided the first study of the phenomenon of *template ageing* over a time span of two years, supported by the recordings collected in the Mars 500 isolation study. We showed the application of the HV-MOTS framework to audio problems through generic audio samples classification and sound morphology.

We showed how this knowledge gained through broader applications can be put to use in the field of musical orchestration. We introduced a new orchestration system based on a algorithm that use the MOTS framework and that rely on a entropic segmentation procedure. We showed that this new algorithm outperforms the previous approaches for computer-aided orchestration. We then presented other artistic applications of the MOTS framework.

On a more epistemological level, we can see that the analysis of musical problematics offers a very powerful framework to study wider scientific topics. As we have discussed along this study and in our directions of future work, the inherent temporal nature of music writing exhibit very stimulating problematics. Through the connections that exist between the micro-level of spectral properties and the macro-time of musical discourse, rise the question of *temporal granularity* and even further the idea of a *dense temporal scales continuum*.

# Part VII

# APPENDIX

# A HV-MOTS DATASETS DESCRIPTION

### ARABIC-DIGIT (UCI)

**Field** Spoken digit recognition

**Summary** This dataset contains the recordings of 10 spoken arabic digit by 88 speakers. Each digit is repeated 10 times for each speaker The data comes from 44 males and 44 females native Arabic speakers

**Features** Sound files have been analyzed to obtain 13 Mel-Frequency Cepstrum Coefficients (MFCCs) time series.

**Classes** Spoken arabic digits ($[0-9]$)

**Samples** 8800 samples (10 digits x 10 repetitions x 88 speakers)

**Sampling** Computation from sound files at 11025 Hz sampling rate, 16 bits with Hamming window. A pre-emphasis filter was applied to the original signal.

**Source** This dataset is part of the UCI repository [126] and is extensively described in [158, 157].

**Results** A Vector Quantization (VQ) with a Maximum Weight Spanning Tree (MWST) leads to a final mean classification accuracy of 93.12% over every classes with single classes results varying from 85.55% to 99.00%

### ARTIFICIAL CHARACTERS (UCI)

**Field** Character recognition

**Summary** This dataset has been artificially generated by using first order theory to describe the structure of ten capital letters of English alphabet. Each instance is described by a set of segments which imitate the way an automatic program would segment an image.

**Features** Each segment is represented by $X$ and $Y$ values for the starting and ending points.

**Classes** 10-classes problem representing the capital letters $A$, $C$, $D$, $E$, $F$, $G$, $H$, $L$, $P$ and $R$

**Samples** 6000 samples (600 for each class)

**Source** This dataset is part of the UCI repository [126] and is described in Botta et al. [52]

**Results** A Genetic Algorithm coupled with an histogram local optimization leads to a recognition rate of 98.68%

### AUSTRALIAN-SIGNS (UCI)

Field Sign recognition

Summary This dataset consists of samples of Auslan (Australian Sign Language) signs that were recorded with multiple sensors on a powered glove. Examples of 95 signs were collected from five signers with a total of 6650 sign samples

Features The glove recorded 10 different time series feature for each sign as the *x*, *y* and *z* position of the hand, the *roll*, *pitch* and *yaw* of the hand orientation and the bending for *thumb, forefinger, index, ring* and *little* fingers.

Classes Each sign represents a different class which amounts to 95 different classes.

Samples 6650 samples (varies for each class)

Source This dataset is part of the UCI dataset repository [126] and is extensively described in [186].

Results For the low quality set (Nintendo Power-glove), best results are obtained by a Hidden Markov Model (HMM) classifier with 71.2% classification accuracy [187]

### AUSTRALIAN-SIGNS-HQ (UCI)

Field Sign recognition

Summary This dataset is a highest quality version of the Australian-signs dataset. However, recordings were made with a single signer. It contains 27 examples of each of the same 95 signs for a total of 2565 signs collected from a native signer using high quality position trackers on both hands.

Features The same feature set is used, however this time both hands were recorded simultaneously.

Samples 2565 samples (27 for each class)

Articles This dataset is part of the UCI dataset repository [126] and is described in [187].

Results For this dataset, best results are obtained by the Naive Segmentation based on TClass algorithm with a 94.5% accuracy [187].

### BCIIII-03A-GRAZ

Field Brain-Computer Interface

Summary This dataset is composed of EEG recordings of cued motor imagery with 4 classes (*left hand, right hand, foot, tongue*) from 3 subjects (ranging from quite good to fair performance). Performance is measured using the kappa-coefficient.

Features 60 EEG channels (1-50Hz)

Classes 4-class problem between *left hand*, *right hand*, *foot* and *tongue*

Samples 840 samples (70 trials per subject for each class)

| | |
|---|---|
| Sampling | 250Hz sampling rate |
| Source | This dataset is part of the 2003 BCI Competition 2003 Blankertz et al. [48]. |
| Results | A multi-class CSP based on Fisher ratios obtained a kappa coefficient of 0.7926. |

### BCIIV-01-BERLIN

| | |
|---|---|
| Field | Brain-Computer Interface |
| Summary | Motor imagery for an uncued EEG classifier application (for *hand* and *foot*); evaluation data is a continuous EEG which also contains periods of idle state |
| Features | 64 EEG channels (0.05-200Hz) |
| Classes | 3 classes (*hand*, *foot* and *idle* state) |
| Samples | 1400 samples (200 trials per subject) |
| Sampling | 1000Hz sampling rate |
| Source | This dataset is part of the BCI Competition IV Blankertz et al. [49]. The performance measure is the mean squared error with respect to the target vector |
| Results | The output of the competition shows that a clustering procedure applied on a Principal Components Analysis (PCA) allows to obtain a MSE of 0.382. |

### BCIIV-03-FREIBURG

| | |
|---|---|
| Field | Brain-Computer Interface |
| Summary | Hand movement directions in MEG. The data set contains directionally modulated low-frequency MEG activity that was recorded while subjects performed wrist movements in four different directions. |
| Features | 10 MEG channels (filtered to 0.5-100Hz) located above the motor areas |
| Classes | 4 classes (*forward*, *backward*, *left* and *right* wrist movements) |
| Samples | 480 samples (120 for each classes with 60 trials per subject) |
| Technical | The trials were cut to contain data from 0.4 s before to 0.6 s after movement onset and the signals were band pass filtered (0.5 to 100 Hz) and resampled at 400 Hz. |
| Source | This dataset is part of the BCI Competition IV Blankertz et al. [49]. |
| Results | The final obtained accuracy is 46.9% over the 2 subjects with 59.5% for the first and 34.3% for the second. Feature extraction and reduction is then fed to a Genetic Algorithm (GA) which decide the features to use for classification with a combination of linear Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) |

### BIOMAG-2010

| | |
|---|---|
| Field | EEG analysis |
| Summary | The goal of this dataset is to detect whether subjects are attending to the left or right visual field on each trial based on the MagnetoEncephaloGram (MEG) of the subjects. |
| Features | 274 MEG channels recorded independently |
| Classes | 2-class problem (between *left* and *right* visual field). |
| Samples | 780 samples |
| Source | This dataset is described in Van Gerven and Jensen [359] |
| Results | The results reported in Van Gerven and Jensen [359] shows that a Support Vector Machine (SVM) algorithm provide an accuracy of up to 69% correctly classified trials. |

### CHALLENGE-2011

| | |
|---|---|
| Field | Cardiology |
| Summary | This dataset is composed of 12-lead ECG recordings, with the aim of providing useful feedback on the quality of the ECG signals by classifying them depending on their quality (from poor to excellent). |
| Features | This dataset is composed of 12 leads (*I, II, II, aVR, aVL, aVF, V1, V2* and *V3*) |
| Classes | 2-class task between *acceptable* and *unacceptable* ECG recording depending on their qualities. |
| Samples | 2000 samples (1000 for each class) |
| Technical | The leads are recorded simultaneously for a minimum of 10 seconds; each lead is sampled at 500 Hz with a 16-bit resolution. |
| Source | This dataset is part of the *PhysioBank* archive [140] posted as the 2011 challenge. |
| Results | The challenge was intended to see the selectivity and specificity of algorithms. The best system reports a 85.9% accuracy Xia et al. [384] using the spectrum radius of a matrix of regularity. |

### CHARACTER-TRAJECTORIES (UCI)

| | |
|---|---|
| Field | Character recognition |
| Summary | This dataset is composed of labelled samples of pen tip trajectories recorded for individual characters writing. All samples are from the same writer, for the purposes of primitive extraction. Only characters with a single pen-down segment were considered |
| Features | The dataset is made of three time series for each instance, which represents the $x$ and $y$ position of the pen and the *pen tip force*. |

Classes 20-class problem over different characters

Samples 2858 samples (varies for each class)

Technical Recordings have been made at 200 Hz sampling rate. Data has been numerically differentiated and Gaussian smoothed, with a sigma value of 2.

Source This dataset is part of the UCI repository [126] and is described in Williams et al. [378, 379].

Results Reported classification accuracy of 87% in Zafar et al. [401] and 93.67% in Perina et al. [276] using a Gaussian Mixture Model (GMM) with a Hidden Markov Model (HMM).

### DACHSTEIN

Domain High altitude medicine

Summary Each instance represents EEG and ECG data for one cardiac cycle that were acquired at 900 m and at 2700m altitude. The subject performed a reaction time task. The data shows the influence of the loss of oxygen on event-related desynchronization (ERD) and event-related synchronization (ERS) and heart rate variability.

Features The recordings covers 2 channels of EEG (C3 and C4) and the ECG recordings.

Classes 2-class problem between *900m* and *2700m* recordings

Samples 698 samples (324 at 900m and 374 at 2700m)

Technical 256 Hz sampling rate with a 1 μV calibration

Source This dataset is described in Guger et al. [151] where analysis is performed to show the influence of loss of oxygen

Results - No classification accuracy reported -

### EEG-ALCOHOLISM (UCI)

Field Medical analysis

Summary This data comes from a large study to examine EEG correlates of genetic predisposition to alcoholism. It contains measurements from 64 electrodes placed on subject's scalps.

Features Each recording is composed of *64 EEG* electrodes time series

Classes 2-class problem between *alcoholic* and *control* subjects

Samples 650 samples

Technical Data was recorded at 256 Hz sampling rate

Source This dataset is part of the UCI repository [126] and is described in [406].

Results The results provided in Zhong and Ghosh [408] show that a Multivariate HMM allows a classification accuracy of 90.5%. However, this result is obtained with only 10 pre-selected measurements. The true classification accuracy is 78.5%

FORTE

Domain Climatology (lightning prediction)

Summary This dataset is split into three different problems depending on the number of classes they contain. Each dataset is aimed at predicting the type of lightning observed through recordings of the power density.

Features Ground density of Electro-Magnetic Power (EMP) recording

Samples 121 instances in each dataset.

Source The dataset is extensively described in[100].

Results Classification resultsof 77.5% accuracy are reported in Bernecker et al. [41] using a Shared-Nearest-Neighbors (SNN) algorithm with the Longest Common SubSequence (LCSS) distance.

**Forte-2**

Classes 2-class problem where distinction should be made between *cloud-to-ground* and *intra-cloud* lightning.

**Forte-6**

Classes 6-class problem where distinction should be made between *CG* (Positive-Initial Return Stroke), *SR* (Subsequent Negative Return Stroke), *IR* (Negative Initial Return Stroke), *I* (Impulsive Event), *I2* (Impulsive Event Pair) and *KM* (Gradual Intra-Cloud Stroke) lightning events

GAITPDB

Field Medical analysis

Summary This database contains measures of gait from patients with idiopathic Parkinson Disease (PD) (mean age: 66.3 years; 63% men), and healthy control subject (mean age: 66.3 years; 55% men). A disturbed gait is a common, debilitating symptom of PD; patients with severe gait disturbances are prone to falls and may lose their functional independence. The database includes the vertical ground reaction force records of subjects as they walked at their usual, self-selected pace for approximately 2 minutes on level ground.

Features This dataset was recorded using 8 sensors under each foot, which gives the *vertical ground reaction force* for each foot as well as the *total force* under each foot. This amounts to 18 time series features.

Classes 2-class problem between *healthy* and *parkinson* subjects based on gait disturbances.

Samples 306 samples (214 parkinson and 92 healthy subjects)

Technical Recordings were made at 100 Hz sampling rate

Source This dataset is part of the PhysioBank database [140] and is described in Frenkel-Toledo et al. [128]

Results Results introduced in Lee and Lim [219] shows that Neural Network with weighted fuzzy membership functions on Wavelet Transform coefficients allow to obtain a 77.33% classification accuracy.

### HANDWRITTEN (UCI)

Field Character recognition

Summary This dataset contains 8235 online handwritten assamese characters. The "online" process involves capturing the data in real-time while the characters are written on a digitizing tablet with an electronic pen. This dataset was collected from 45 writers, each of which contributed 183 recordings

Features The acquisition program records the handwriting as a stream of $X$ and $Y$ coordinate points using the appropriate pen position sensor along with the pen-up and pen-down switching. No pressure level was recorded.

Classes This is a 183-classes problem. Each writer contributed 52 basic characters, 10 numerals and 121 assamese conjunct consonants.

Samples 8235 samples (45 for each class).

Source This dataset is part of the UCI repository [126]

Results A classification accuracy of 92% with HMM to 96% with SVM is reported but **only for digits** (so only a 10-classes problem)

### IONOSPHERE (UCI)

Field Radar analysis

Summary This dataset was collected by a radar system in Goose Bay, Labrador. The targets were free electrons in the ionosphere. "Good" radar returns are those showing evidence of some type of structure in the ionosphere. "Bad" returns are those that do not; their signals pass through the ionosphere.

Features There were 17 pulse numbers for the Goose Bay system. Instances in this database are described by 2 attributes per pulse number, corresponding to the complex values returned by the function resulting from the complex electromagnetic signal which amounts to 34 features.

Classes 2-class binary task between *good* and *bad* radar returns.

Samples 358 samples

Technical The recording system consists of a phased array of 16 high-frequency antennas with a total transmitted power of 6.4 kilowatts. Received signals are processed using an autocorrelation function whose arguments are the time of a pulse and its number.

Source: This dataset is part of the UCI repository [126] and is described in Sigillito et al. [329]

Results: The best results is a 94.2 % classification accuracy reported in Eggermont et al. [114] using a Genetic Programming (GP) approach.

### JAPANESE-VOWELS (UCI)

Field: Speaker identification

Summary: This dataset covers the topic of speaker identification for nine male speakers which uttered two Japanese vowels successively.

Features: Each instance is represented by 12 LPC cepstrum coefficients time series.

Classes: 9-class problem representing each speaker

Samples: 640 samples (varies for each class)

Technical: 10 kHz sampling rate analyzed with a 25.6ms window size and a 6.4ms hop size.

Source: This dataset is part of the UCI repository [126] and is described inKudo et al. [213].

Results: The proposed classifier exhibits a classification accuracy of 94.1%, while a 5-state continuous Hidden Markov Model attained up to 96.2%

### LIBRAS (UCI)

Field: Movement recognition

Summary: The dataset contains 15 classes of 24 instances each. Each class represents to a hand movement type in LIBRAS (oficial brazilian signal language).

Features: In each frame, the centroid pixels of the segmented objects (the hands) are found. These successive points compose the discrete version of the curve which is represented by the $x$ and $y$ position time series. All curves are normalized in the unitary space.

Classes: 15-class problem for *swings*, *arcs*, *circle*, *lines*, *zigzag*, *waves*, *curves* and *tremble* movements

Samples: 360 samples (24 for each class).

Technical: In the video pre-processing, a time normalization is carried out selecting 45 frames from each video, by following an uniform distribution.

Source: This dataset is part of the UCI repository [126], described originaly in Dias et al. [105] and later used in Schliebs et al. [318]

Results: An evolving Spiking Neural Network (eSNN) allows to obtain a 88.59% classification accuracy.

### PEN-CHARS-35 (UCI)

| | |
|---|---|
| Field | Character recognition |
| Summary | This dataset is composed of upper and lowercase characters, digits, and other spanich characters, for a total of 62 different characters collected from 11 different writers which performed 2 repetitions. |
| Features | Each character is represented by a sequence of segments summarized by their *X* and *Y coordinates* for each point. |
| Classes | 62-classes problem. |
| Samples | 1364 samples (2 repetitions for 11 writers for each class) |
| Source | This dataset is part of the UCI repository [126] and is described in Prat et al. [286] |
| Results | Classification accuracy is shown to be up to 89.15% if using a DTW matching algorithm on segment-based representation with a NN-rule. |

### PEN-CHARS-97 (UCI)

| | |
|---|---|
| Field | Character recognition |
| Summary | This dataset is composed of ASCII and non-ASCII characters which amount to a total of 97 different characters collected from 60 different writers which performed 2 repetitions. |
| Features | Each character is represented by a sequence of segments summarized by their *X* and *Y coordinates* for each point. |
| Classes | 97-classes problem. |
| Samples | 11640 samples (2 repetitions for 60 writers for each class) |
| Source | This dataset is part of the UCI repository [126] and is described in Castro-Bleda et al. [78] |
| Results | Classification accuracy is presented in Castro-Bleda et al. [78] of 91.8% classification accuracy if using a template matching algorithm with a NN-rule however this results is *only for a restricted set of 62 characters.* |

### PERSON-ACTIVITY (UCI)

| | |
|---|---|
| Field | Movement analysis |
| Summary | People used for recording of the data were wearing four tags (ankle left, ankle right, belt and chest). Each instance is a localization data for one of the tags. The goal of this dataset was to dect falls only but we also test the accuracy in classifying all movement classes. |
| Features | The four tags worn by subjects sent the $x$, $y$ and $z$ position for *left* and *right ankle*, *chest* and *belt* which amounts to 12 distinct features. |

Classes — This dataset features instances of different postures and actions with *walking*, *falling*, *lying*, *lying down*, *sitting*, *sitting down*, *standing up from lying*, *on all fours*, *sitting on the ground*, *standing up from sitting* and *standing up from sitting on the ground*.

Samples — 164860 samples (varies for each class)

Source — This dataset is part of the UCI repository [126] and is described in [189]

Results — Classification accuracy is shown to be 72% for machine learning agents, 88% for expert-knowledge agents and 91.3% for meta-prediction agents. However, this results are for detecting falls only.

### PHYSICAL-ACTION (UCI)

Field — Movement analysis

Summary — Three male and one female subjects (age 25 to 30), who have experienced aggression in scenarios such as physical fighting, took part in the experiment. Throughout 20 individual experiments, each subject had to perform ten normal and ten aggressive activities.

Features — The data acquisition process involved eight skin-surface electrodes placed on the upper arms (biceps and triceps), and upper legs (thighs and hamstrings), which corresponds to eight input time series for all muscle channels (ch1-8). Each time series contains around 10000 samples (15 actions per experimental session for each subject). The electrodes were placed on *right bicep* (C1), *right tricep* (C2), *left bicep* (C3), *left tricep* (C4), *right thigh* (C5), *right hamstring* (C6), *left thigh* (C7) and *left hamstring* (C8)

Classes — 2-class problem between *aggressive* and *normal* actions

Samples — 80 samples (4 for each class)

Technical — The subjects' performance has been recorded by the Delsys EMG apparatus, interfacing human activity with myoelectrical contractions.

Source — This dataset is part of the UCI repository [126]

Results — A Genetic Programming (GP) algorithm allows to obtain 73.3% classification accuracy Theodoridis and Hu [353]. However these results are for six of the actions only.

### PTBDB

Field — Cardiology

Summary — The ECGs were collected from healthy volunteers and patients with different heart diseases

Features — Each record includes 15 simultaneously measured cardiac signals. The conventional 12 ECG-leads (*i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6*) together with the 3 Frank lead ECGs (*vx, vy, vz*).

Classes 9 separate class representing healthy patients and patients suffering different heart conditions. The classes are therefore *healthy controls*, *myocardial infarction*, *cardiomyopathy*, *bundle branch block*, *dysrhythmia*, *myocardial hypertrophy*, *valvular heart disease*, *myocarditis* and *miscellaneous*.

Samples 2750 instances from 549 records performed on 290 subjects (aged 17 to 87, mean 57.2; 209 men, mean age 55.5, and 81 women, mean age 61.6; ages were not recorded for 1 female and 14 male subjects). Each subject is represented by one to five records.

Technical Each signal is digitized at 1000 samples per second, with 16 bit resolution over a range of ± 16.384 mV. with 0.5 μV/LSB (2000 A/D units per mV).

Source This dataset is part of the PhysioBank database [140] and is described in [53].

Results The results provided in Gudmundsson et al. [150] shows that a Random Forest (RF) classifier provide a 75.1% classification accuracy

**Ptdb-1**

Technical This dataset contains single heart beats for classification

**Ptdb-2**

Technical This dataset contains two heart beats in each instance

**Ptdb-5**

Technical This dataset contains five heart beats in each instance

ROBOT-FAILURES (UCI)

Field Robotics

Summary This dataset contains force and torque measurements on a robot after failure detection. Each failure is characterized by 15 force/torque samples collected at regular time intervals

Features The robots measure a set of $x$, $y$ and $z$ *force position* as well as a $x$, $y$ and $z$ *torque* measured after failure, which amounts to a total of 6 individual time series features.

Classes This dataset is divided into five sub-sets, each of them defining a different learning problem

Technical Each failure instance is characterized in terms of 15 force/torque samples collected at regular time intervals starting immediately after failure detection. A total observation window of 315ms is used for each failure instance.

Source This dataset is part of the UCI repository [126] and described [66].

Results A set of five feature transformation strategies (based on statistical summary features, discrete Fourier transform, etc.) allows to obtain a maximal classification accuracy of 80%.

**Robot-failures-lp1**

Classes Failures in approach to grasp position
24% normal - 19% collision - 18% front collision - 39% obstruction

Samples 88

**Robot-failures-lp2**

Classes Failures in transfer of a part
43% normal - 13% front collision - 15% back collision - 11% collision to the right - 19% collision to the left

Samples 47

**Robot-failures-lp3**

Classes Position of part after a transfer failure
43% normal - 19% slightly moved - 32% moved - 6% lost

Samples 47

**Robot-failures-lp4**

Classes Failures in approach to ungrasp position
21% normal - 62% collision - 18% obstruction

Samples 117

**Robot-failures-lp5**

Classes Failures in motion with part
27% normal - 16% bottom collision - 13% bottom obstruction - 29% collision in part - 16% collision in tool

Samples 164

SLPDB

Field Sleep apnea analysis

Summary The MIT-BIH Polysomnographic Database is a collection of recordings of multiple physiologic signals during sleep. Subjects were monitored in Boston's Beth Israel Hospital Sleep Laboratory for evaluation of chronic obstructive sleep apnea syndrome, and to test the effects of constant positive airway pressure (CPAP), a standard therapeutic intervention that usually prevents or substantially reduces airway obstruction in these subjects.

Features All recordings include an *ECG signal*, an *invasive blood pressure* signal (measured using a catheter in the radial artery), an *EEG signal*, and a r*espiration signal* (in most cases, from a nasal thermistor). Seven-channel recordings also include a *respiratory effort signal* derived by inductance plethysmography, an *EOG signal* and an *EMG signal* (from the chin). Therefore the dataset is divided depending on the number of available features.

Classes Several tags can be applied at the same time for a subject. The first kind depends on the different sleep stage, depending on if the subject is *awake*, in *sleep stage 1, 2, 3, 4* or *REM sleep*. Then phases of apnea are recorded between different types, with *hypopnea*, *obstructive apnea* and *central apnea* that can be *with* or *without arousal*. Final different movements are recorded for *legs* and *arousal*.

Summary 2-class task between *apnea* and *normal* based on 4 features (*ECG, BP, EEG* and *Respiration*)

Samples 4085 samples

Source This dataset is part of the PhysioBank database [140] and described in [175].

Results Between 83.24% to 88.97% classification accuracy Bsoul et al. [61] using a Multi-Scale Support Vector Classifier (MS-SVM)

### SONAR

Domain Sonar analysis

Summary This dataset contains the patterns of sonar signals bouncing off either a metal cylinder or rocks. In both case, the angles and conditions varies. The goal is to correctly identify the metal cylinder returns.

Features Set of energy in 60 frequency bands over a certain period of time

Classes 2-class problem between *mine* and *rock* sonar signals.

Samples 208 samples (111 mines and 97 rocks)

Source TThis dataset is part of the UCI repository [126] and is described in **Tan and Dowe [348].**

Results The results reports a 76% classification accuracy by using a Minimum Message Length (MML) Oblique Tree.

### SYNEMP

Domain Climatology (lightning prediction)

Summary The classification tasks are aimed at studying varying speed of leading edge for different classes of lightning

Features Synthetic density of Electro-Magnetic Power (EMP) recording

Classes 2-class problem between *slow* and *fast* leading edge

Samples 20000 samples (10000 for each class)

Source This dataset is described in [184].

Results - No classification results reported -

### VFDB

Field Cardiology

Summary This database includes 22 half-hour ECG recordings of subjects who experienced episodes of sustained ventricular tachycardia, ventricular flutter, and ventricular fibrillation.

Features Two *ECG* recordings based on separate electrodes.

Classes 15-class problem between different ventricular fibrillations comprising *atrial fibrillation*, *asystole, ventricular bigeminy, first degree heart block, high grade ventricular ectopic activity, normal sinus rhythm, nodal ("AV junctional") rhythm, noise, pacemaker (paced rhythm), sinus bradycardia, supraventricular tachyarrhythmia, ventricular escape rhythm, ventricular fibrillation, ventricular flutter* and *ventricular tachycardia.*

Samples 600 samples (40 per class)

Source This dataset is part of the PhysioBank database [140] and is described in [147]

Results A band-pass digital filtration and ECG peak detection algorithm allows to obtain a 91.5% classification algorithm in Krasteva and Jekova [212].

### VICON-PHYSICAL (UCI)

Field Physiological analysis

Summary This dataset includes 10 normal and 10 aggressive physical actions based on various human activities. The data have been collected by 10 subjects using the Vicon 3D tracker.

Features The $x$, $y$ and $z$ positions define the 3D position of each marker in space for *left* and *right wrist*, *elbow*, *ankle* and *knee* which amounts to a total of 26 time series features for each action.

Samples 2000 instances (10 for each action)

Technical The duration of each action was approximately 10 seconds per subject, which corresponds to a time series of 3000 samples, with sampling frequency of 300Hz.

Source This dataset is part of the UCI repository [126] and is described in [353].

Results 95.4% classification accuracy is reported using a Dynamic Neural Network, however it is only applied on 9 actions out of the 20 classes.

Classes 2-class problem between *aggressive* and *normal* actions

### WALL-ROBOT

Domain Robotics

Summary This dataset contains ultrasound readings for a wall-following task for robotic navigations.

| | |
|---|---|
| Features | 24 ultrasounds readings and 4 minimum sensor readings. |
| Classes | 4-class problem between *forward*, *slight-right*, *sharp-right* and *slight-left* movements |
| Samples | 5460 samples (1365 for each class) |
| Technical | Sensor readings are sampled at a rate of 9 samples per second. |
| Source | This dataset is part of the UCI repository [126] and is described in **Freire et al. [127].** |
| Results | The results exhibit a classification accuracy of 95.58% if using a Polynomial kernel SVM. |

# B HEART SOUNDS BIOMETRY

## B.1 CARDIAC AUSCULTATION

*Auscultation*

Stethoscopic auscultation still plays today a leading role in medical diagnosis. Developed with the work of Laennec in 1816 [215], the auscultation with stethoscopes has remained almost unchanged since. The late 1940s has witnessed the arrival of a new graphical method for cardiac analysis: the *PhonoCardioGram* (PCG) which associated with the recording of pulsatile phenomena allowed a precise diagnosis of cardiac pathologies. Despite its successful development, this technique has been gradually replaced since the 1970s by ultrasound imaging obtained by continuous and pulsed Doppler. However, the required echocardiographic equipment is still extremely expensive. Furthermore, the extensive training required for recording and interpretation of ultrasounds makes it a highly specialized technique. Auscultation, meanwhile, has the merit of simplicity and can be performed anywhere at any time. The last years have also witnessed the emergence of electronic stethoscopes directly equipped with noise reduction and enhancement filters. Auscultation has kept his sense today as a unique diagnostic tool for the detection of cardiac anomalies. PCG has considerably expanded its scope with the advent of graphic reproduction and signal processing techniques. For decades, it has still not be supplented as a basis for accurate diagnosis of heart disease and is now systematically associated with the pulsatile carotid recording, or even direct recording of impulse on the chest wall.

Auscultation is usually performed through different beaches, defined by their position on the chest or abdomen. Auscultation beaches can allow doctors to hear several slightly varyiating cardiac cycles which enhance diagnosis by a better localisation of the anomaly. Cardiac auscultation beaches are named according to their relative positions to the heart valves. Some diseases of the thoracic aorta can lead to use a back auscultation, but most data is collected on the anterior chest wall at the aortic, pulmonary, mitral and tricuspid beaches. It is recognized that 4G2 (fourth left intercostal space 2 cm from the midline) is a beach which often provides a good summary of cardiac cycles.

*Pathologies*

Several cardiovascular diseases can alter the mecanisms of the human heart and consequently the sounds that it produces. We list here what kind of modifications can occur in PCG recordings and briefly discuss the related pathologies.

SUPERNUMARY NOISES In addition to S1 and S2, a third (S3) and fourth (S4) heart sounds may also be audible. These sounds occur in the diastole phase and are the result of cardiac insufficiencies. S3 usually succeeds S2 (from 80 to 140ms afterwards) and can be traced to a massive spontaneous filling of the left ventricule. This sound is always the sign of a ventricular pathology. S4 preceeds S1 and is related to the occurence of atrial contraction caused by an active ventricular filling. This sound sometimes appear in healthy adolescents but is distinctive of a pathology after 35 years of age.

systolic noises Sometimes, an isolated high-frequency systolic noise occurs shortly after S1. This noise is caused by a deficient opening of the aortic valve and is also known as *systolic ejection click*. Another opening snap can appear shortly after S2 with the prolapse of the mitral valve. These kind of sounds are always distinctive features of heart deficiencies.

diastolic noises These sounds similar to a snap, originates in the mitral opening with rigidified valves. Therefore it appears approximately 80ms after S2. Supernumerary diastolic noises are usually associated with valvular or pericardial disease. It is sometimes followed by a roll which exhibit a left ventricular filling through a restricted orifice with high atrial pressure. A high-frequency click can also occur at the beginning of the diastole because of a constricted pericardium.

sounds splits The usual S1 and S2 sounds can sometimes be heard as duplicated, which leads to "split sounds". These are the most common anomalies and their occurence indicates a deficiency in the right heart that causes a loss of the usual synchronism between the left and right ventricles.

heart murmurs Heart murmurs are usually long noises that can be heard at different times of the cardiac cycle. Murmurs are a consequence of a change in the streaming flow inside the heart. Some of these can be louder depending on their origin. Ejection murmurs are related to transvalvular flow and are very common situations where they are termed as "innocents" as they are not pathological indicators. Their classification is therefore frequently based on their timing of occurence which can be diastolic, systolic or even continuous.

*PCG Biometry*

The idea of using heart sounds (*PhonoCardioGram* (PCG)) as a biometric feature has has first been introduced by Phua et al. [280]. They proposed to study heart sounds through a Short-Term Discrete Fourier Transform (STDFT) in order to obtain their spectral decomposition. The decomposition is then processed with a bandpass filter in order to remove frequency bins outside the useful cardiac sounds information. Finally, dimension reduction and spike removing allows to filter out noise and artifacts from the spectral information. In order to match the identification templates, they compared the use of Vector-Quantization (VQ) and Gaussian Mixture Model (GMM) with different sets of features (Mel-Frequency Cepstral Coefficients (MFCC) and Linear Frequency Band Cepstra (LFBC)). They extended their work in [281] with a larger database of 128 people and showed that the LFBC features computed on 256ms frames with a 4-component GMM classification scheme was the most accurate system. However, it should be noted that these works use the *same recordings* that are split in enrollment and identification sequences, therefore it cannot accounts for the inherent variability between different recordings conditions. Beritelli and Serrano [35] independently proposed a biometric identification system by using frequency analysis of heart sound recordings. In this paper, the authors first segmented the cardiac recordings into the two main sound components. They then applied the Chirp $z$-Transform (CZT) on segmented sounds to obtain the spectral information from each cardiac cycle. In order to assess the identity, feature vectors are compared to stored templates by using the Euclidean distance. They further developed this system in [36] where they used Linear Frequency Cepstrum Coefficients (LFCC) and a feature specifically designed for heart sounds called the First-to-Second Ratio (FSR) which defines the proportionality of average powers between the first (*S1*) and second (*S2*) heart sounds.

For the past years these ideas has simply been extended by different approaches. Tran et al. [355] investigated the use of various feature sets by exploring temporal, spectral, harmonic and rythmic features. They further applied a feature selection algorithm in order to classify heart sounds with automatically selected features. El-Bendary et al. [117] extended the spectral decomposition approach by using the Discrete Wavelet Transform (DWT) on which they extracted correlation and cepstral features. In order to classify heart sounds they compared Mean Square Error (MSE) and k-Nearest Neighbors (kNN) classifiers and showed that the latter improved classification accuracy on their database. Jasper and Othman [183] also used the DWT with Daubechies wavelet, limited to sub-bands between 30 and 140Hz. After pre-processing and filtering the signal, they showed the superiority of the Shannon energy envelogram on their dataset.

Other approaches have also been undertaken. Fatemian et al. [121] tried to combine ECG and PCG-based approaches for identification in order to improve classification accuracy. They processed the heart sounds with a STFT and Mel-filterbank before applying a Linear Discriminant Analysis (LDA) which allows to obtain low-dimensional vectors on which Euclidean distance can be applied. Beritelli and Spadaccini [37] proposed a statistical approach for PCG-based biometry where they use a Gaussian Mixture Model with Universal Background Model (GMM-UBM) recognition technique. This technique appears to improve accuracy over previous structural approaches.

## B.2 S-FEATURES

In order to precisely separate the properties of heart sounds, we need to use a spectral decomposition that can fit and enhance their unique characteristics. Several signal processing tools have been developed over the past years, such as the Short-Time Fourier Transform (STFT). However heart sounds have an extremely narrow useful bandwith. Furthermore, most of their energy lie in very low frequency ranges, below the resolution power of usual decompositions. The Stockwell transform was originally developed for analyzing geophysical data [338]. This time-frequency distribution is defined as a generalization of both the Short-time Fourier transform and the Continuous Wavelet Transform (CWT) but overcomes some of their limitations. There are several ways to express the S transform. As stated by [338], the S-Transform can be defined as a CWT with a gaussian mother wavelet multiplied by a phase factor. Therefore, the S-Transform of a function $h(t)$ is defined as

$$S_x(t,f) = \int_{-\infty}^{\infty} h(\tau) \frac{|f|}{\sqrt{2\pi}} e^{-\frac{(t-\tau)^2 f^2}{2}} e^{-i2\pi f\tau} d\tau \tag{B.1}$$

Compared to classical spectral decomposition, the S-Transform provides a frequency dependent resolution which leads to a finer definition. In our case, what is more relevant is that the S-Transform (as the CWT) can give access to an extremely fine resolution even at very low frequencies (below 50Hz) which is the most useful bandwidth of heart sound signals. Therefore it allows a better distinction of spectral components.

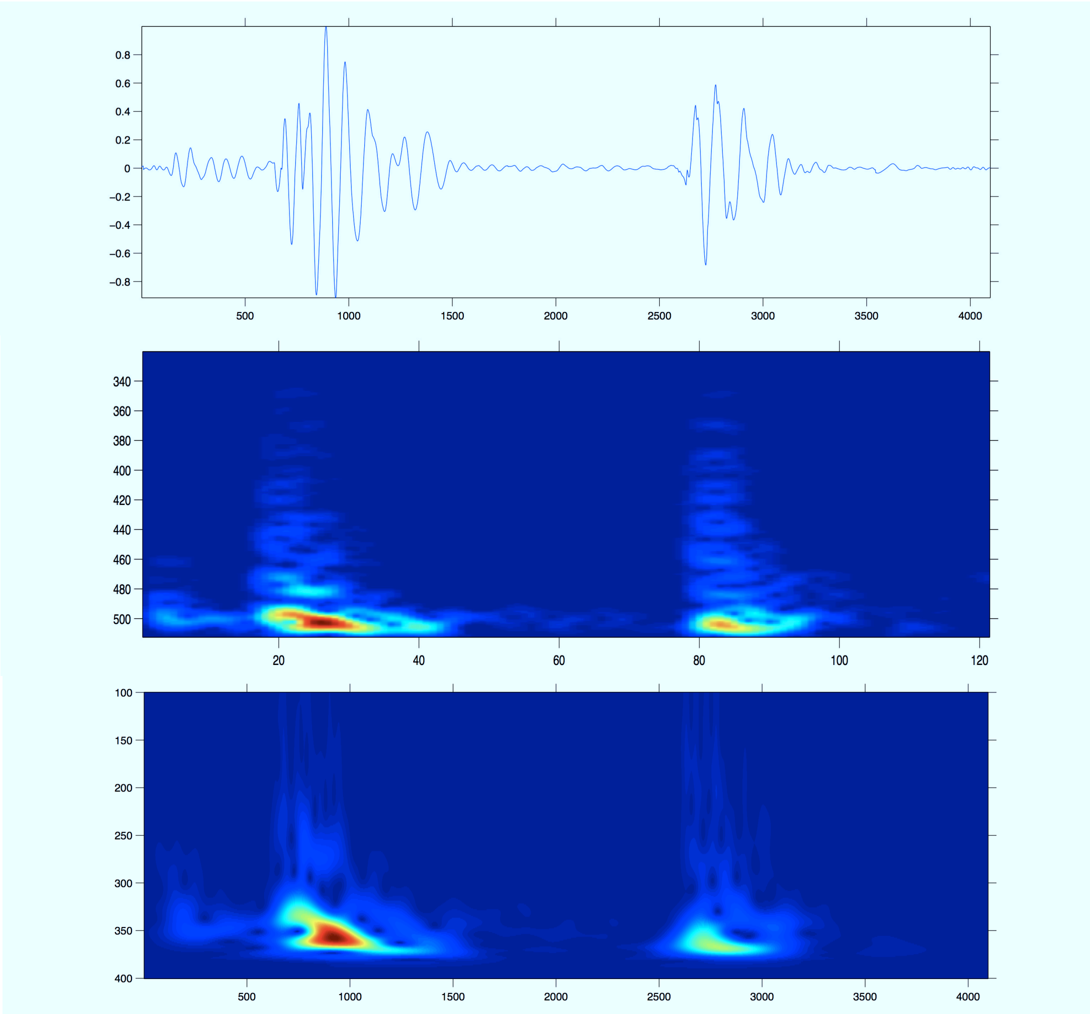



Figure 79: Comparing the resolution power of the FFT to the S-Transform for a single cardiac cycle.

Furthermore, unlike the CWT, modulation sinusoids are fixed with respect to the time axis. This localizes dilations and translations and thus provides the same temporal resolution for each frequency bin. A fast S-Transform algorithm was proposed [60] which strongly reduces its computational complexity and therefore makes it usable in real-life applications.

Based on the computation of the S-Transform $S_x(t, f)$, we can access precise information on the frequency distributions. However, we still need to extract high-level information that could put forward the uniqueness of each heart. To that end, we developed a specifically-tailored set of features inspired by the work in musical analysis and content-based audio retrieval. We call this set of high-level information the *S-Features*.

### B.2.1 *Statistical moments*

We can study the evolution of the shape of a distribution by computing its statistical moments. That way, we can gain insights on the distribution of the energy in various frequency bins over time.

S-CENTROID The first moment is the geometric center (*centroid*) of the distribution, it allows to measure which frequency defines the central position of the energy distribution

$$S_{centroid}(t) = \int f.S_x(t,f)\,df \tag{B.2}$$

S-SPREAD The S-Spread is based on the second statistical moment (or *variance*) which exhibits the dispersion of energy distribution over the frequency bins

$$S_{spread}(t) = \int (f - S_{centroid}(t))^2\, S_x(t,f)\,df \tag{B.3}$$

S-SKEWNESS The S-Skewness is based on the third statistical moment which measures the symmetry of the distribution. More precisely, the skewness exhibit the lack of symmetry for a distribution, where a positive value indicates that the distribution have more values larger than the mean. Convertly, a negatively skewed distribution has more values lower than the mean. A symmetrical distribution has a skewness of zero.

$$S_{Skewness}(t) = \frac{\int (f - S_{centroid}(t))^3\, S_x(t,f)\,df}{\left(\sqrt{S_{spread}(t)}\right)^3} \tag{B.4}$$

S-KURTOSIS The S-Kurtosis is based on the fourth statistical moment and is computed by using its corresponding definition

$$S_{kurtosis}(t) = \frac{\int (f - S_{centroid}(t))^4\, S_x(t,f)\,df}{S_{spread}(t)^2} \tag{B.5}$$

The kurtosis summarizes whether the distribution is more peaked or flat than the normal distribution.

### B.2.2 *Energy distribution*

S-BRIGHTNESS The S-Brightness allows to study the distribution of energy over the frequency range. The idea is to fix a cut-off frequency $f_c$ and measure the percentage of energy above that threshold, therefore yielding the temporal function

$$S_{Brightness}(t) = \frac{\sum_{f_i > f_c} S_x(t,f_i)}{\sum_i S_x(t,f_i)} \tag{B.6}$$

In our implementation we tested cutoff frequencies in the set $\{65, 100, 135, 170, 200, 250\}$Hz. We also combined these functions to provide a multi-dimensional per band descriptor called S-BrightnessBands.

S-ENTROPY The Shannon entropy, extensively used in information theory, gives a generic description of shapes. It especially allows to determine whether a distribution contains predominant peaks or not. In order to compute the S-Entropy, we use the definition of relative entropy which is independent of the sequence length

$$S_{entropy}(t) = -\frac{\sum_{i=1}^{N} S_x(t, f_i) \log_b S_x(t, f_i)}{\log_b(n)} \tag{B.7}$$

S-FLATNESS The S-Flatness is a measure which indicates whether the distribution of energy over frequencies is smooth or if spikes are present in a spectral frame. It is computed by dividing the geometric mean by the arithmetic mean of each frame of the S-Transform

$$S_{flatness}(t) = \frac{\sqrt[N]{\prod_{i=1}^{N} S_x(t, f_i)}}{\left(\frac{\sum_{i=1}^{N} S_x(t, f_i)}{N}\right)} \tag{B.8}$$

S-RMS The Root-Mean-Square (RMS) energy of each frame of the S-Transform can be computed by taking the root average of the square of each frequency bin

$$S_{RMS}(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_x(t, f_i))^2} \tag{B.9}$$

S-ROLLOFF The S-Rolloff allows to study the temporal evolution of energy concentration. The idea is to find the frequency $f_c$ such that a certain fraction of the total energy $thresh_e$ is contained below that frequency.

$$S_{rolloff}^{thresh_e}(t) = f_c \mid \sum_{i=1}^{f_c} S_x(t, f_i) < t_e \tag{B.10}$$

It is interesting to note that $S_{rolloff}^{0.5}(t)$ gives an approximation of the S-Centroid. In our implementation we compute the S-Rolloff for $\{65, 75, 85, 95\}$ % of energy and also provide the multidimensional S-RolloffBands.

S-FLUX The S-Flux describes the difference between two successive frames of the S-Transform. It allows to obtain a measure of "novelty" in the S-Transform by showing whether two successive frames are similar or not

$$S_{flux}(t_i) = \sqrt{\sum_{j=1}^{N}\left(S_x\left(t_{i+1}, f_j\right) - S_x\left(t_i, f_j\right)\right)^2} \tag{B.11}$$

### B.2.3 *Peaks distribution*

The following descriptors are all based on the spectral peaks found in the S-Transform. Therefore, we first apply a peak tracking algorithm to the spectral frames. We then consider that we have access to two ordered list $\mathcal{P}_i^f$ and $\mathcal{P}_i^a$ which contains respectively the frequencies and amplitudes of the prohiminent peaks found in the spectrum.

**S-IRREGULARITY** The S-Irregularity allows to exhibit the degree of variation of contiguous peaks found in the spectrum. It is therefore obtained by computing the mean difference of amplitude between each successive pair of peaks.

$$S_{Irregularity}(t) = \frac{\sum_{i=1}^{N_p-1}\left(\mathcal{P}_i^a - \mathcal{P}_{i+1}^a\right)^2}{\sum_{i=1}^{N_p}\left(\mathcal{P}_i^a\right)^2} \tag{B.12}$$

**S-ROUGNESS** The S-Roughness allows to study the closeness of prohiminent peaks in each frame of the S-Transform. Originally based on the auditory concept of beating phenomenon, we simplified its computation by using the average distance in frequency between all pair of peaks.

$$S_{Roughness}(t) = \frac{\sum_{i=1}^{N_p-1}\sum_{j=i}^{N_p-1}\left(\mathcal{P}_i^f - \mathcal{P}_j^f\right)^2}{\sum_{i=1}^{N_p}\left(\mathcal{P}_i^f\right)^2} \tag{B.13}$$

## B.3 EXTENDED HEARTS BIOMETRY ANALYSIS

We provide in this section an extended analysis of the hearts biometric system. As several parts of the proposed system may have an influence on the overall performance, we study each of them separately. These comparisons are performed following the same order as our algorithmic workflow (cf. Figure 46).

### B.3.1 *Pre-processing*

The first step in the treatment of heart beats is to prepare the signal in order to remove its eventual impurities. As every audio signals, heart recordings are subject to *noise* and *spikes* that we pre-process thanks to different filtering techniques.

*Despiking filter*

The despiking filter allows to remove the *spikes* and *jitters* that might appear due to defects in the recording system. Figure 80 exhibits the influence of the despiking filter through the ROC curves of *with* and *without* use of the filter. As we can see, the use of the despiking algorithm strongly enhance the performances of the algorithm. However, this improvement seems to be more noticeable on the ROC curve (and therefore on the overall accuracy of the system) than on the rank identification rates. Therefore, the despiking filter enhance the overall performances but less on a per-person basis. This can be explained by the fact that beats which exhibit strong outliers (spikes) are "fixed" by the filter. Without filtering, these beats provoke wide error anomalies (their scores are extremely lower than normal beats). Hence, without these, the *global* score analysis exhibit the improvement. Therefore, this seems to indicate that the despiking filter is an efficient pre-processing step. However, this also indicates that the set of beats which requires such analysis seems to be of small cardinality.

*Wavelet denoising*

The wavelet denoising allows to handle the presence of background noise in the signal by removing its eventual presence and thus improve the *Signal-to-Noise Ratio* (SNR).

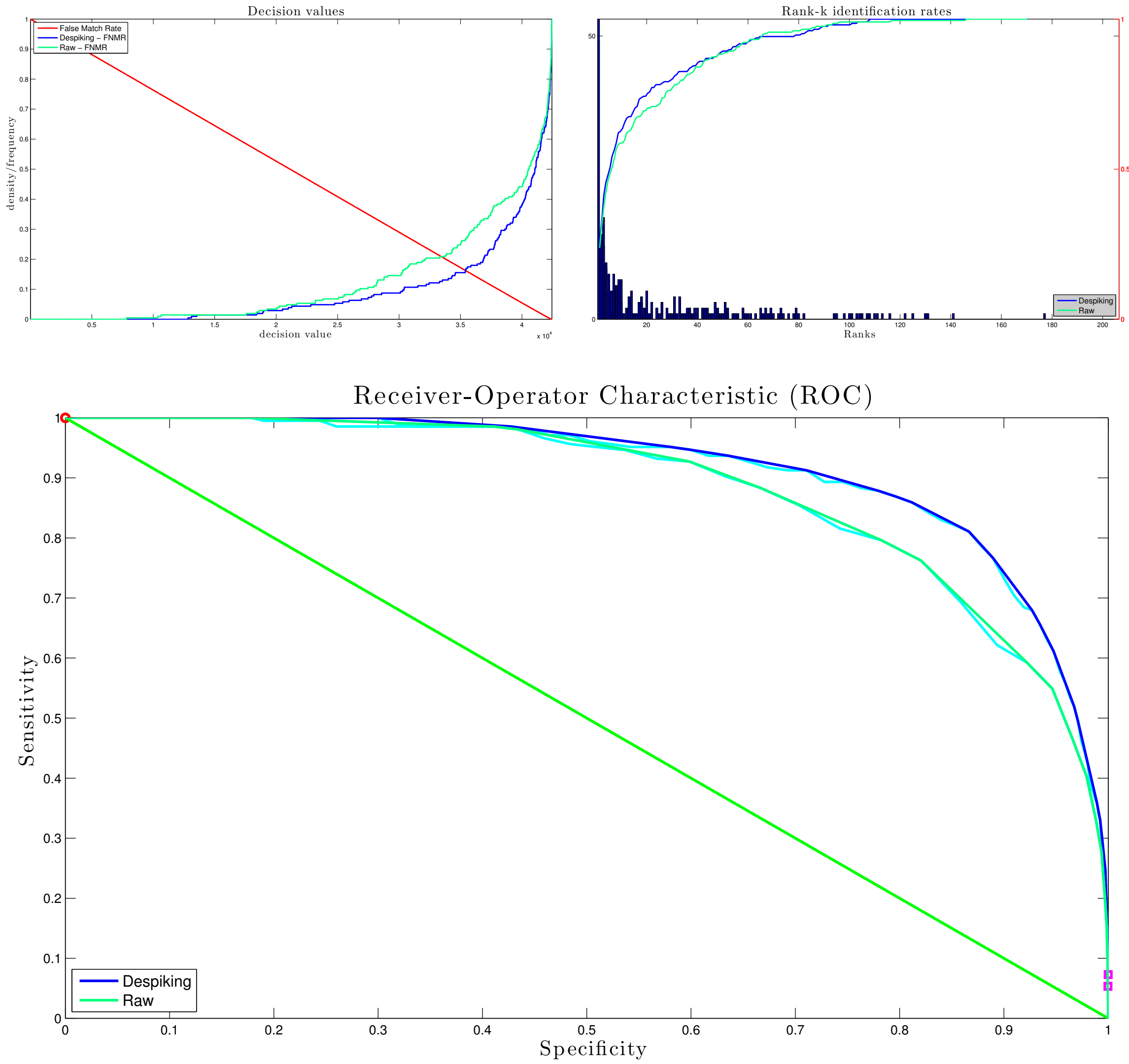



Figure 80: Influence of the despiking filter exhibited through the ROC curves of *with* and *without* use of the filter.

When using such denoising algorithm, the choice of the wavelet is crucial to its success. Figure 81 exhibits the influence of the wavelet choice through the ROC curves of using the *Coiflets* or *Daubechies* wavelets. As we can see on the ROC curves, the use of the denoising filter also enhance the overall accuracy of the system. It seems that the use of either wavelets provide the same improvement (the two curves are almost merged). However, compared to the despiking filter, the wavelet processing seems to also strongly impact the per-person scores (as seen in the rank identification rates). It seems logical as this time, it is not a set of *single beats* that exhibit defects (throughout every person), but rather a set of *complete recordings* only for a few person. These recordings being more noisy than others, can provoke the complete scores of one individual to exhibit lower accuracy. Therefore, when these recordings are fixed, the corresponding person obtains a better ranking.

### B.3.2 *Segmentation*

The segmentation procedure is applied after all pre-processing has been performed, in order to obtain a set of coherent heart beats from the entire recording. It is therefore a crucial part for the success of the biometric system (it can be related to the *template extraction* procedure). However, the segmentation algorithm is based on several choices for its parameters. Therefore, Figure 82 shows the influence of all segmentation parameters through the resulting ROC curves. Distinction is made between the choice of *positive* or *negative* difference in spectral flow, sizes of the analysis window of 0.7, 0.85 or 1 seconds with a number of partials between 0 and 2. We can see here that there is a whole span of different results depending on the settings of the segmentation procedure. First, the best results are obtained with the *positive* spectral flow for an analysis window of *0.85s* and *2* partials connectivity used (these are the parameters used for the final analysis of results). Therefore, the corresponding curves are merged in the final figure. The widest disparities appears for the size of the analysis window. This parameter widely influence the final results, as it seems that a slightly smaller or wider size of window can dramatically change the accuracy. The two other parameter also seem to strongly influence the results. Therefore, a correct segmentation appears to be the most important step of the whole heart biometric system.

### B.3.3 *Beat Selection*

After the segmentation procedure, a set of heart beats is obtained for each individual. However, this set may often contain some erroneous beats recordings, due to widely variating recordings. Therefore, we perform a selection of segmented beats in order to improve the homogeneity of the resulting set. This selection can be performed either on *energy deviation* (we select the heart beats which are almost of same total energy) and *shape deviation* (we select the set based on minimizing the inter-beats distance in energetic time series). Therefore, figure 83 exhibits the influence of each of these selection procedures based on their resulting ROC curves.

### B.3.4 *Time series comparison*

Once all the heart beats are segmented, the temporal evolution of their features are compared thanks to different time series distances. We try to see the influence and

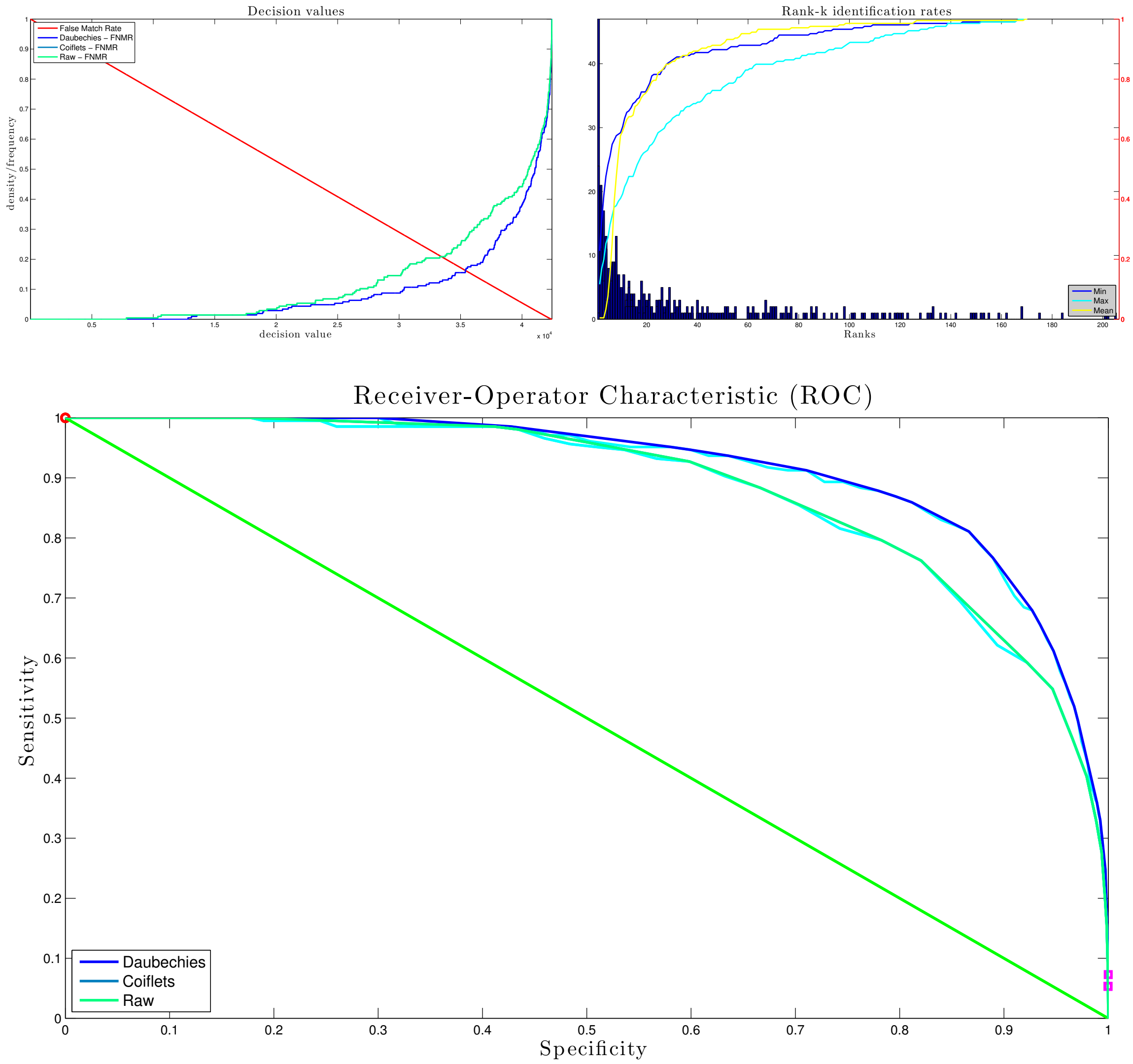



Figure 81: Influence of the wavelet denoising exhibited through the ROC curves of *Coiflets* and *Daubechies* wavelets.

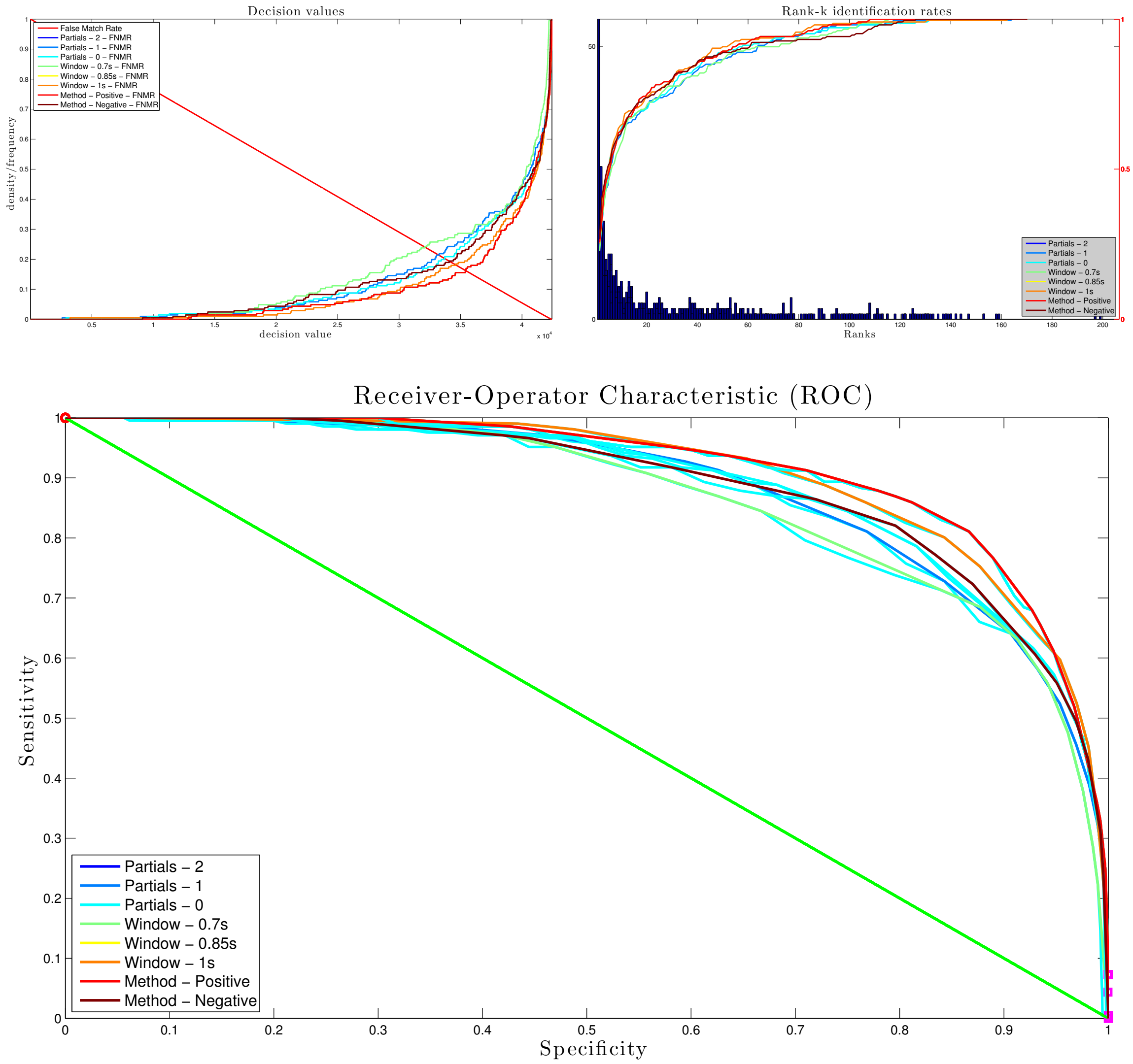



Figure 82: Influence of the segmentation parameters exhibited through the ROC curves of spectral differences (F+ or F-), window size and number of partials.

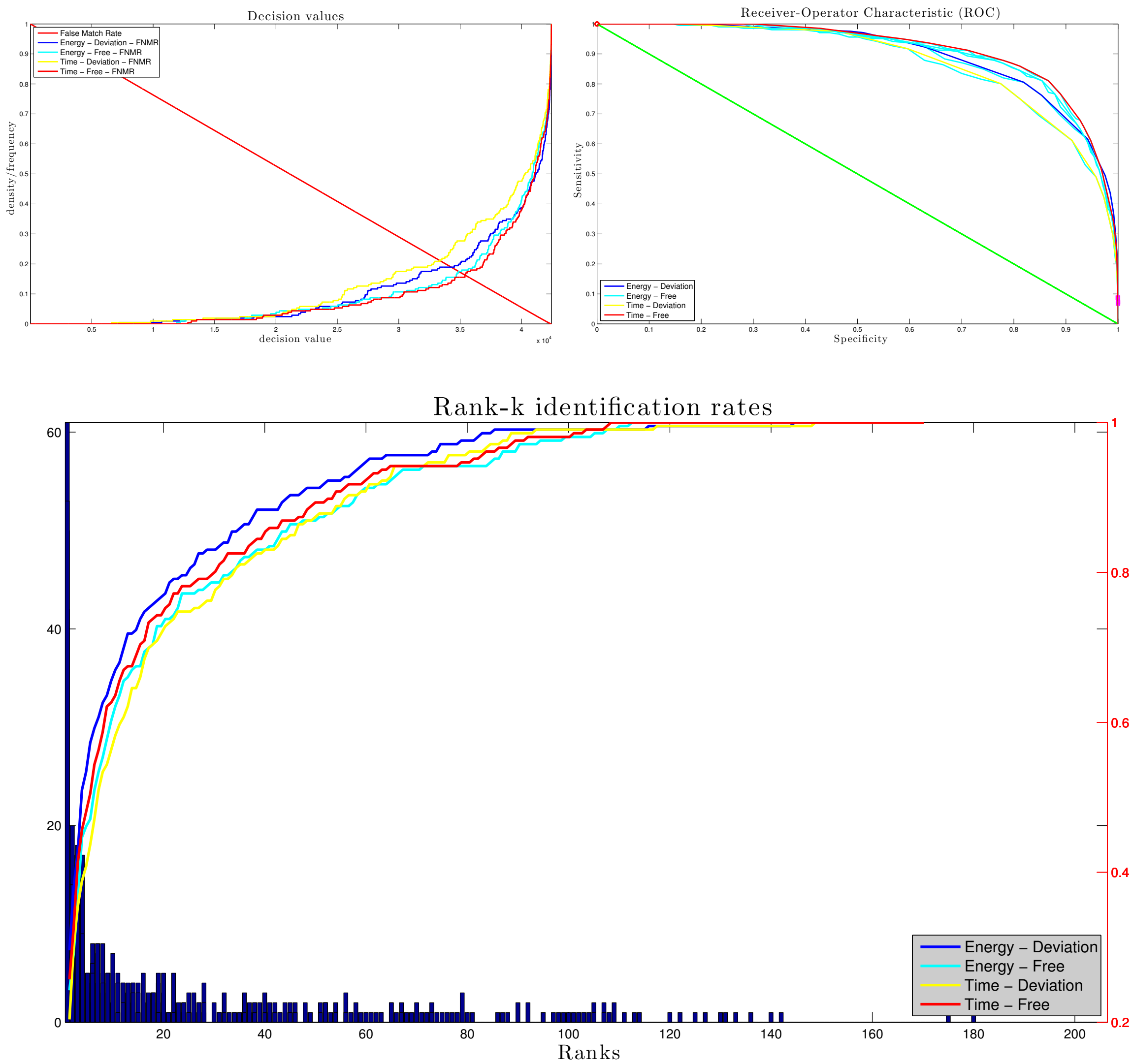



Figure 83: Influence of the beat selection exhibited through the ROC curves of *energy deviation* and *shape deviation*.

differences between the use of DTW with different maximum warping windows and simply using the Euclidean distance on down-sampled time series.

*Warping window*

When using the DTW distance on time series, it is possible to constrain the maximum warping authorized between two series. This parameter allows to control the *reach* to which two points might be compared. Figure 84 shows the influence of the size of the warping window exhibited through the ROC curves of *2*, *5*, *10* and *20*% of authorized warping reach.

*Resampling*

Another solution to compare the time series is to use the Euclidean distance. This distance is often claimed to lack the temporal flexibility required for precise matching. However, our large scale study on various datasets shows that the use of Euclidean distance on strongly down-sampled series often leads to better classification results than using the DTW. Figure 85 shows the influence of the size of the resampling factors for comparing the time series exhibited through the ROC curves of *128, 64, 32, 16* and *8* resampling points.

### B.3.5 *Decision influence*

We study here the influence of the decisions alternatives on the overall performances. This decision relies on the way the final similarity score is computed for one individual. First the type of *decision rule* decides how the score of multiple heart beats should be merged together. The number of heart beats to be merged is decided by the *size of testing set*. Each heartbeat is compared to the distribution in the database depending on the *size of training set*.

*Type of decision rule*

When a recording is input to the system, it provides several heartbeats to compare to the templates in the database. Therefore, we obtain a set of scores between each heartbeat and each class in the database. In order to make a final decision, we have therefore different possibilities of merging these scores. Figure 86 shows the influence of the type of *decision rule* exhibited through the ROC curves of the *mean*, *min* and *max* rules.

*Size of testing set*

The previous comparison made the assumption that all heart beats obtained from the segmentation and selection are used for the final score. However, as our selection procedure allows to rank each heart beat based on its *energy* and *shape* deviations, we try to see if we can obtain better results by constraining the maximum number of heart beats to use in the final scoring. Figure 87 shows the influence of the size of *testing set* exhibited through the ROC curves of different set cardinalities.

*Size of training set*

Based on the same idea as the previous comparison, we provide a comparison based on the size of the *training set*. Therefore, when computing the similarity scores, we reduce

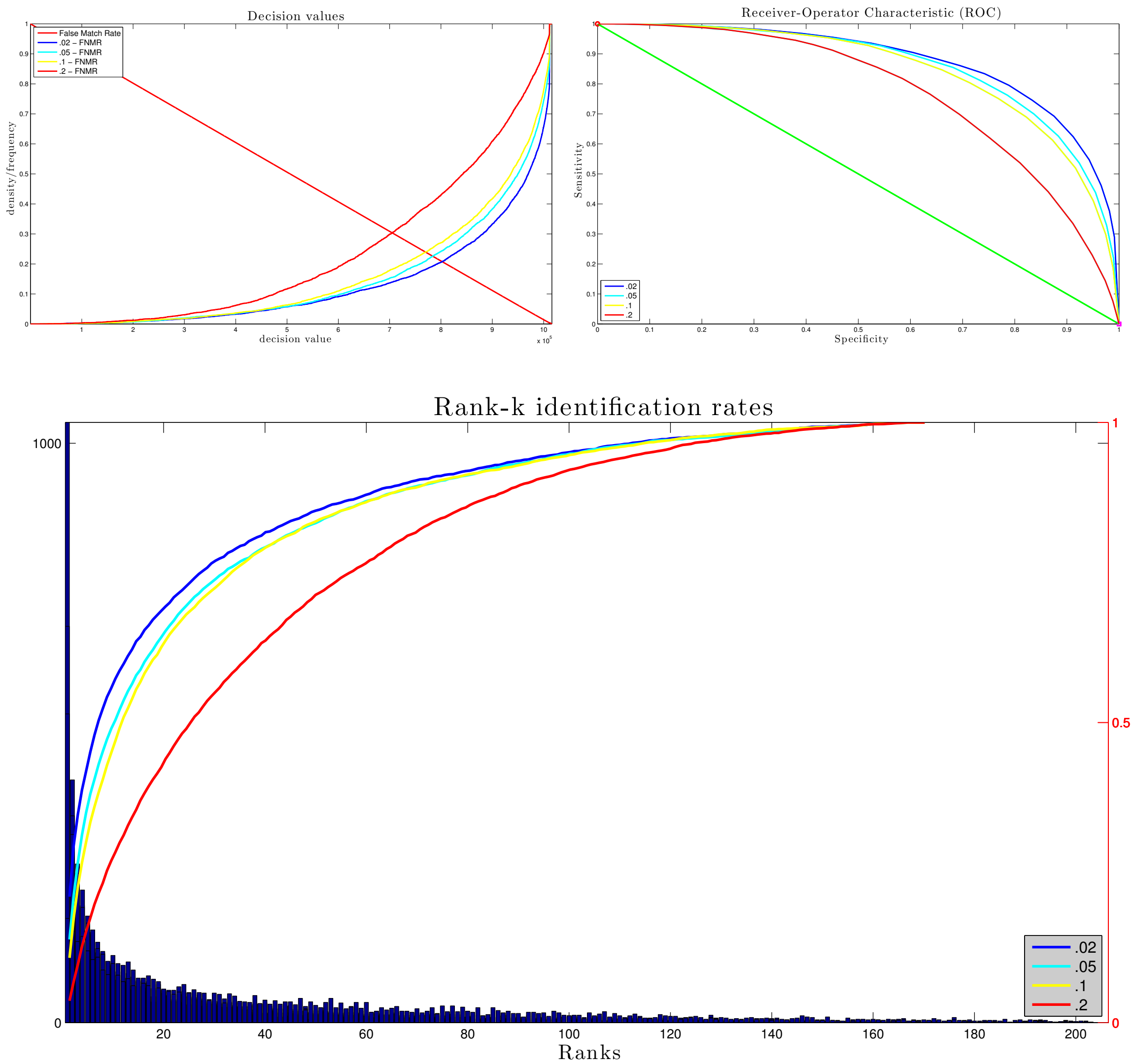



Figure 84: Influence of the size of the warping window for comparing the time series with the DTW distance measure exhibited through the ROC curves of 2, 5, 10 and 20% of authorized warping reach.

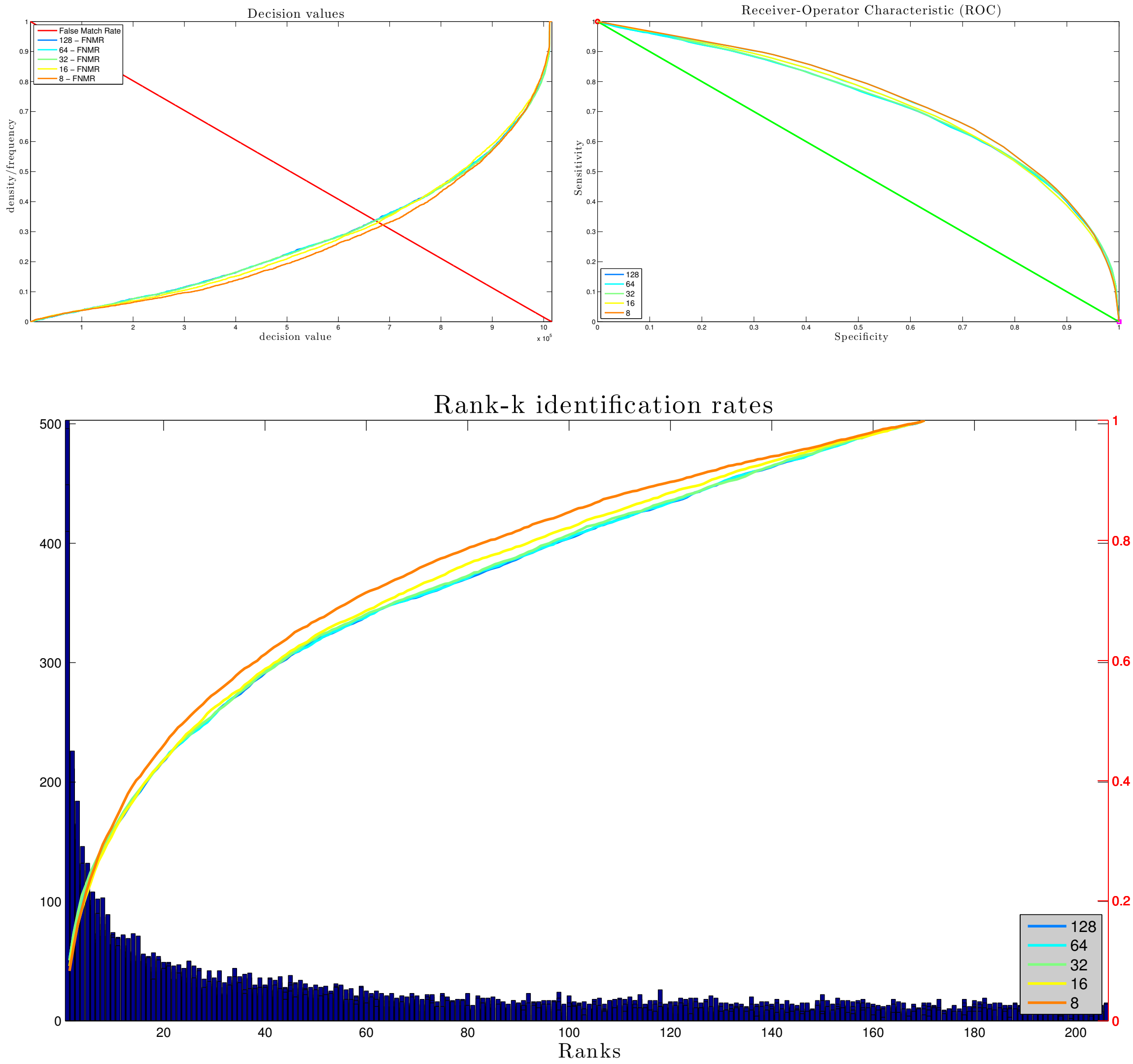



Figure 85: Influence of the size of the resampling factors for comparing the time series exhibited through the ROC curves of *128*, *64*, *32*, *16* and *8* resampling points.

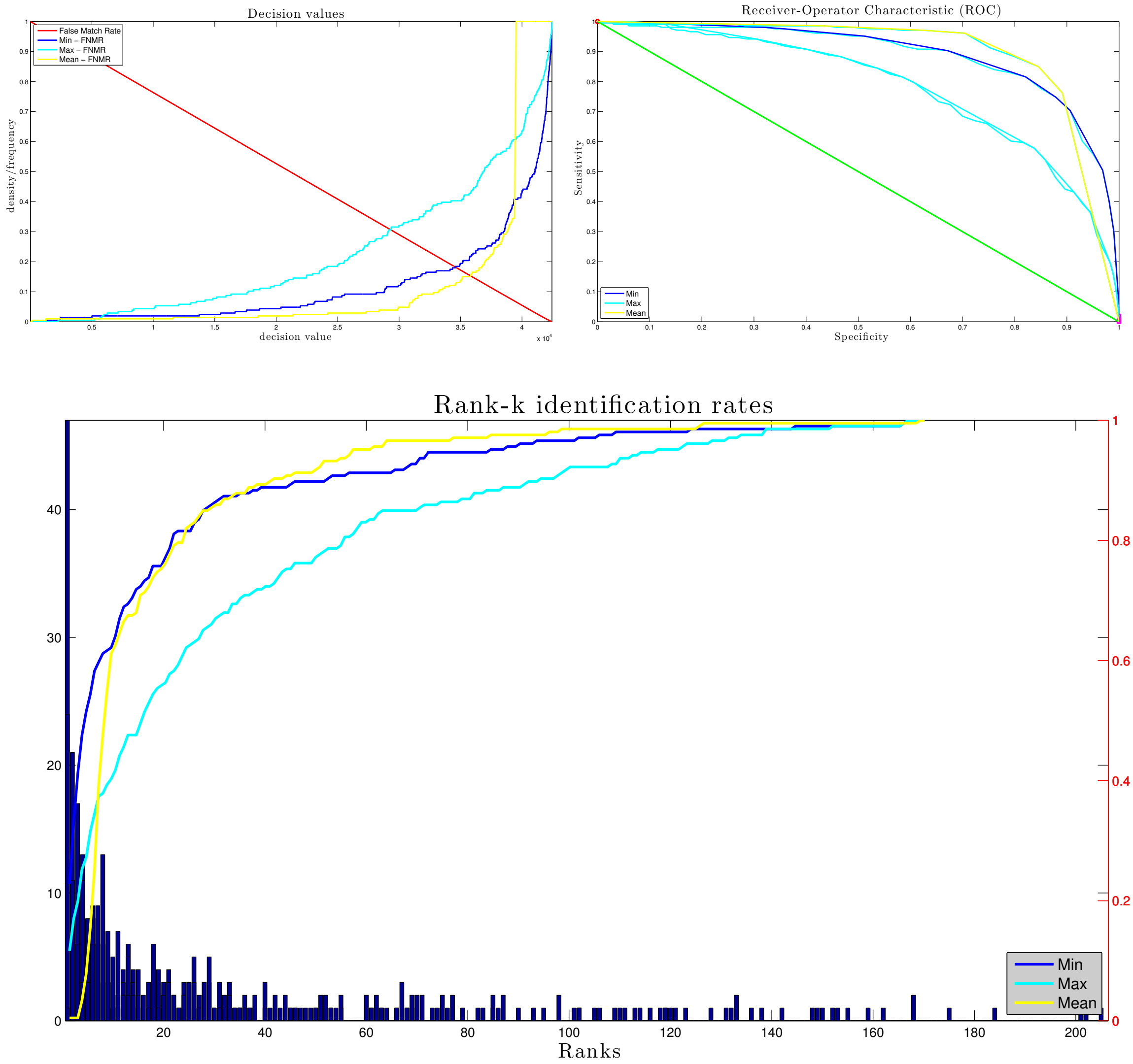



Figure 86: Influence of the type of *decision rule* exhibited through the ROC curves of the *mean*, *min* and *max* rules.

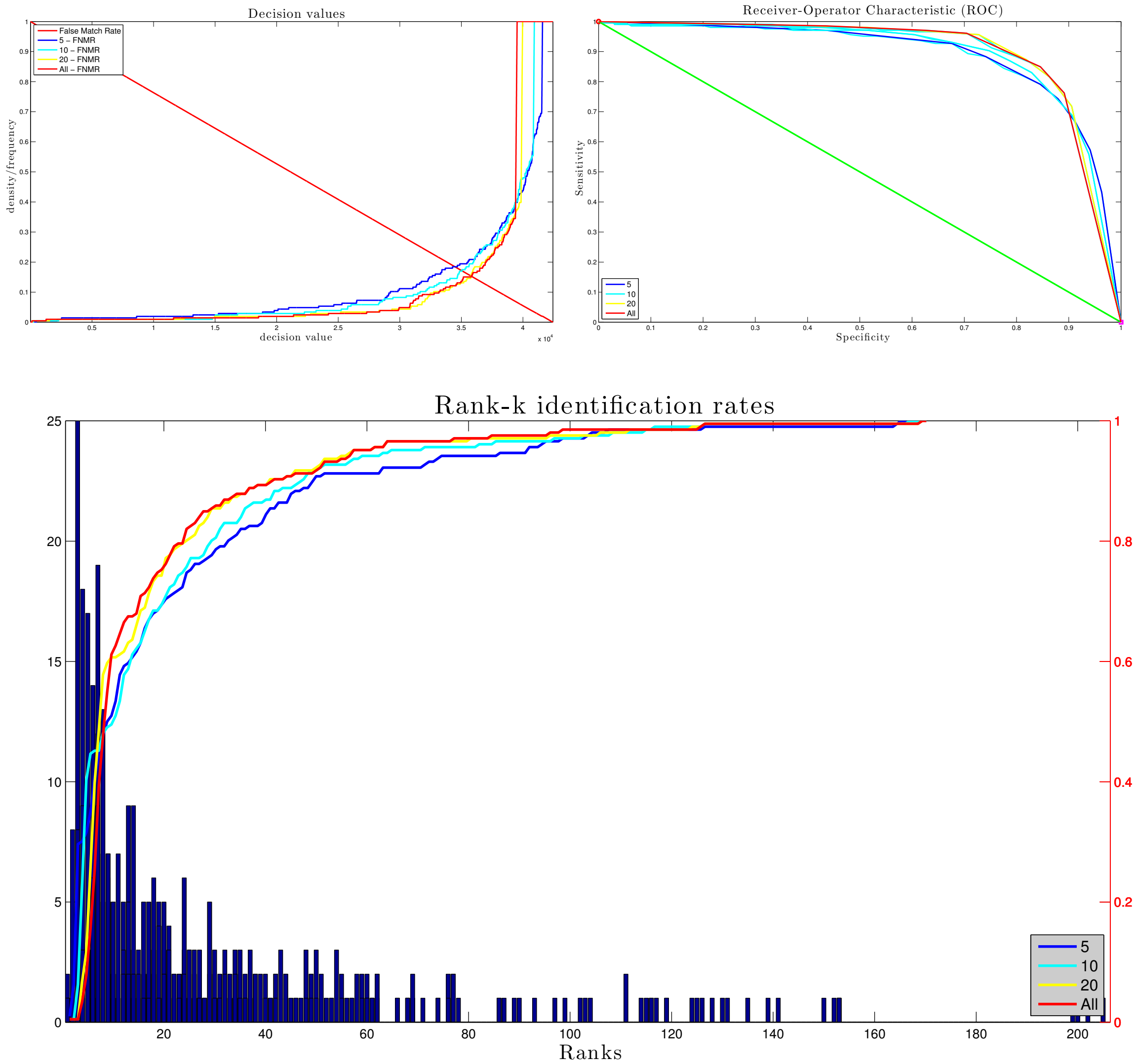



Figure 87: Influence of the size of *testing set* exhibited through the ROC curves of different set cardinalities.

the size of the template database for each individual and use only a restricted set of templates. Figure 88 shows the influence of the size of *training set* exhibited through the ROC curves of different set cardinalities.

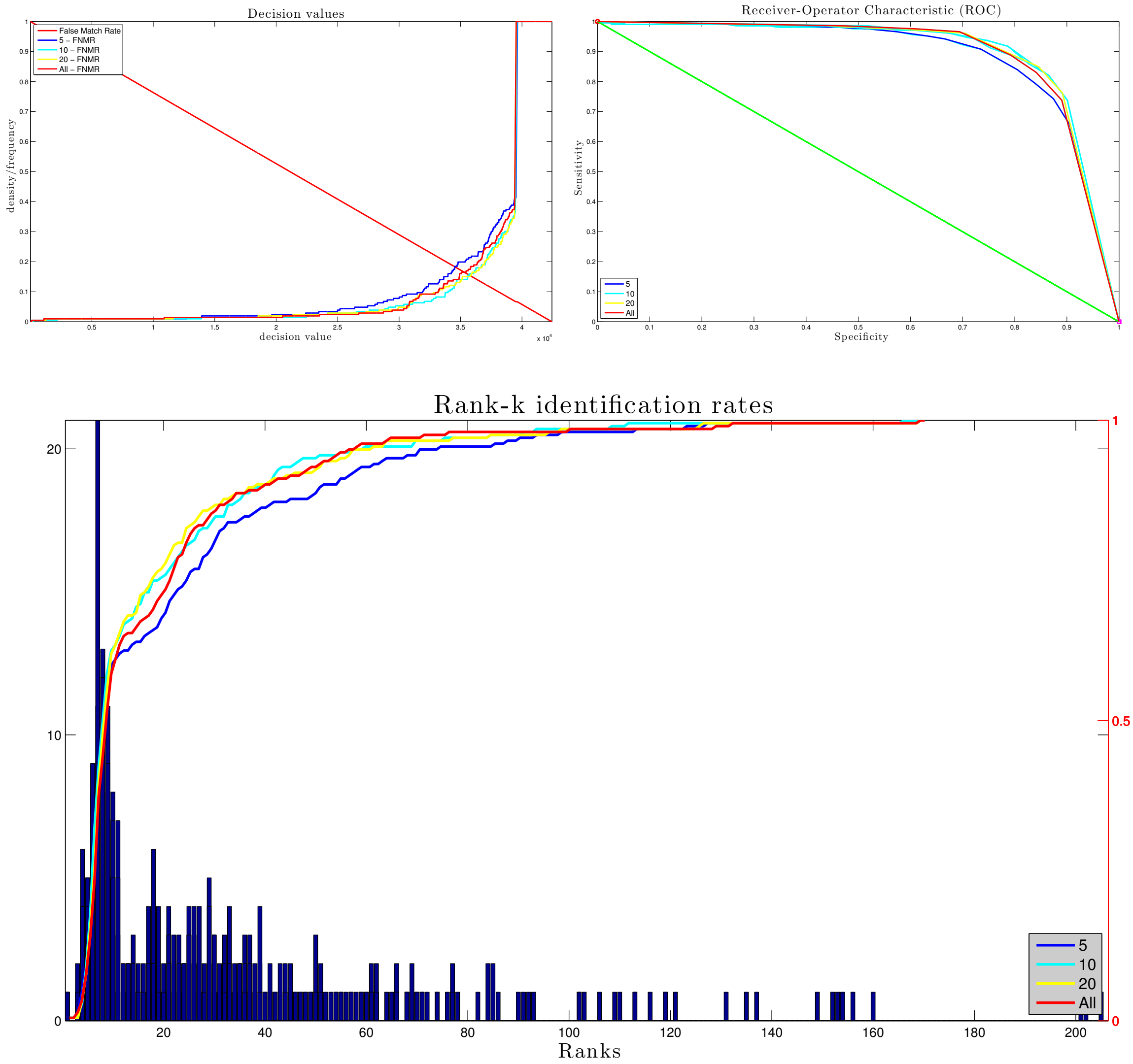



Figure 88: Influence of the size of *training set* exhibited through the ROC curves of different set cardinalities.

# BIBLIOGRAPHY


[1] ISO 9421-11. Ergonomic requirements for office work with visual display terminals - part 11: Guidance on usability. *International Organization of Standardization*, 1995. (Cited on page 79.)

[2] J. Abonyi, B. Fell, S. Nemeth, and P. Arva. Fuzzy clustering based segmentation of time-series. In *Proceedings of the 5th International Symposium on Intelligent Data Analysis, IDA 2003, August 28-30*, pages 275–285, Berlin, Germany, 2003. Springer-Verlag, New York Inc. (Cited on page 31.)

[3] M.F. Abu-Taleb and B. Mareschal. Water resources planning in the middle east: application of the PROMETHEE V multicriteria method. *European Journal of Operational Research*, 81(3):500–511, 1995. ISSN 0377-2217. (Cited on page 54.)

[4] F. Agrafioti, F.M. Bui, and D. Hatzinakos. Medical biometrics: The perils of ignoring time dependency. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*, pages 1–6. IEEE, 2009. (Cited on page 131.)

[5] R. Agrawal, C. Faloutsos, and A.N. Swami. Efficient Similarity Search In Sequence Databases. In *Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms*, pages 69–84, Chicago, Illinois, USA, 1993. Springer. (Cited on pages 26 and 36.)

[6] Rakesh Agrawal, King-Ip Lin, Harpreet S. Sawhney, and Kyuseok Shim. Fast similarity search in the presence of noise, scaling, and translation in time-series databases. In *VLDB '95: Proceedings of the 21th International Conference on Very Large Data Bases*, pages 490–501, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc. ISBN 1-55860-379-4. (Cited on pages 27 and 35.)

[7] N.K. Ahmed, A.F. Atiya, N. El Gayar, H. El-Shishiny, and E. Giza. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5):594–621, 2009. (Cited on page 32.)

[8] T. Ahmed, B. Oreshkin, and M. Coates. Machine learning approaches to network anomaly detection. In *Proceedings of the 2nd USENIX workshop on Tackling computer systems problems with machine learning techniques*, pages 1–6, Cambridge, MA, USA, 2007. USENIX Association. (Cited on page 33.)

[9] J. An, H. Chen, K. Furuse, N. Ohbo, and E. Keogh. Grid-based indexing for large time series databases. *Intelligent Data Engineering and Automated Learning, Lecture Notes in Computer Science*, 1983(1):614–621, 2003. (Cited on page 38.)

[10] C.M. Antunes and A.L. Oliveira. Temporal data mining: An overview. In *KDD Workshop on Temporal Data Mining*, pages 1–13, San Francisco, CA, USA, 2001. (Cited on pages 23, 41, and 44.)

[11] T. Argyros and C. Ermopoulos. Efficient subsequence matching in time series databases under time and amplitude transformations. In *3rd IEEE International Conference on Data Mining*, pages 481–484, 2003. (Cited on page 35.)

[12] I. Assent, R. Krieger, F. Afschari, and T. Seidl. The TS-tree: efficient time series search and retrieval. In *Proceedings of the 11th International Conference on Extending Database Technology*, pages 25–29, 2008. (Cited on page 47.)

[13] I. Assent, M. Wichterich, R. Krieger, H. Kremer, and T. Seidl. Anticipatory DTW for efficient similarity search in time series databases. *Proceedings of the VLDB Endowment*, 2(1):826–837, 2009. (Cited on page 27.)

[14] J. Aßfalg, H.P. Kriegel, P. Kroger, P. Kunath, A. Pryakhin, and M. Renz. Similarity search on time series based on threshold queries. In *Advances in database technology: EDBT 2006: 10th International Conference on Extending Database Technology, March 26-31*, volume 3896, page 276, Munich, Germany, 2006. Springer-Verlag New York Inc. (Cited on page 43.)

[15] J. Aßfalg, H.P. Kriegel, P. Kröger, P. Kunath, A. Pryakhin, and M. Renz. Similarity search in multimedia time series data using amplitude-level features. In *Proceedings of the 14th international conference on Advances in multimedia modeling*, pages 123–133. Springer-Verlag, 2008. (Cited on page 36.)

[16] American Standards Association. *American Standard Acoustical Terminology*. ASA, New York, 1960. (Cited on pages 8 and 10.)

[17] A. Bagnall and G. Janacek. Clustering time series with clipped data. *Machine Learning*, 58(2):151–178, 2005. (Cited on page 28.)

[18] A. Bagnall, C.A. Ratanamahatana, E. Keogh, S. Lonardi, and G. Janacek. A bit level representation for time series data mining with shape based similarity. *Data mining and knowledge discovery*, 13(1):11–40, 2006. (Cited on page 38.)

[19] AJ Bagnall, G. Janacek, B. De la Iglesia, and M. Zhang. Clustering time series from mixture polynomial models with discretised data. In *Proceedings of the 2nd Australasian Data Mining Workshop*, pages 105–120, 2003. (Cited on page 38.)

[20] J. Bai and S. Ng. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317, 2008. (Cited on page 32.)

[21] BR Bakshi and G. Stephanopoulos. Representation of process trends–IV. Induction of real-time patterns from operating data for diagnosis and supervisory control. *Computers & Chemical Engineering*, 18(4):303–332, 1994. (Cited on pages 23 and 29.)

[22] B.R. Bakshi and G. Stephanopoulos. Reasoning in time: Modeling, analysis, and pattern recognition of temporal process trends. *Advances in Chemical Engineering*, 22:485–548, 1995. (Cited on page 38.)

[23] G Ballet, R Borghesi, P Hoffmann, and F Levy. Studio online 3.0 : An internet "killer application" for remote access to ircam sounds and processing tools. In *Actes des Journees Informatique Musicale*, Paris, France, 1999. (Cited on page 68.)

[24] JP Bandera, R. Marfil, A. Bandera, JA Rodríguez, L. Molina-Tanco, and F. Sandoval. Fast gesture recognition based on a two-level representation. *Pattern Recognition Letters*, 30(13):1181–1189, 2009. (Cited on page 38.)

[25] P. Barone, M.F. Carfora, and R. March. Segmentation, Classification and Denoising of a Time Series Field by a Variational Method. *Journal of Mathematical Imaging and Vision*, 34(2):152–164, 2009. (Cited on page 49.)

[26] G.A. Barreto. Time Series Prediction with the Self-Organizing Map: A Review. *Perspectives of neural-symbolic integration*, 77(1):135–158, 2007. (Cited on page 32.)

[27] L. Barrington, A. Chan, D. Turnbull, and G. Lanckriet. Audio information retrieval using semantic similarity. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*, volume 2, pages 722–725, 2007. (Cited on page 71.)

[28] I. Bartolini, P. Ciaccia, and M. Patella. Warp: Accurate retrieval of shapes using phase of fourier descriptors and time warping distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):142–147, 2005. (Cited on page 43.)

[29] R. Bayer and EM McCreight. Organization and maintenance of large ordered indexes. *Acta informatica*, 1(3):173–189, 1972. (Cited on page 46.)

[30] N. Beckmann, H.P. Kriegel, R. Schneider, and B. Seeger. The R*-tree: an efficient and robust access method for points and rectangles. *ACM SIGMOD Record*, 19 (2):322–331, 1990. (Cited on pages 26 and 46.)

[31] N. Belacel. Multicriteria assignment method PROAFTN: Methodology and medical application. *European Journal of Operational Research*, 125(1):175–183, 2000. ISSN 0377-2217. (Cited on pages 50 and 54.)

[32] N. Belacel and M.R. Boulassel. Multicriteria fuzzy assignment method: a useful tool to assist medical diagnosis. *Artificial intelligence in medicine*, 21(1-3):201–207, 2001. ISSN 0933-3657. (Cited on page 54.)

[33] N. Belacel, P. Vincke, JM Scheiff, and MR Boulassel. Acute leukemia diagnosis aid using multicriteria fuzzy assignment methodology. *Computer methods and programs in biomedicine*, 64(2):145, 2001. ISSN 0169-2607. (Cited on page 54.)

[34] S. Berchtold, D.A. Keim, and H.P. Kriegel. The X-tree: An index structure for high-dimensional data. *Readings in multimedia computing and networking*, 4(1): 451–463, 2002. (Cited on page 46.)

[35] F. Beritelli and S. Serrano. Biometric identification based on frequency analysis of cardiac sounds. *Information Forensics and Security, IEEE Transactions on*, 2(3): 596–604, 2007. (Cited on page 203.)

[36] F. Beritelli and A. Spadaccini. Human identity verification based on mel frequency analysis of digital heart sounds. In *Digital Signal Processing, 2009 16th International Conference on*, pages 1–5. IEEE, 2009. (Cited on page 204.)

[37] F. Beritelli and A. Spadaccini. An improved biometric identification system based on heart sounds and gaussian mixture models. In *Biometric Measurements and Systems for Security and Medical Applications (BIOMS), 2010 IEEE Workshop on*, pages 31–35. IEEE, 2010. (Cited on pages 113, 119, 127, and 204.)

[38] P. Berkhin. A survey of clustering data mining techniques. *Grouping Multidimensional Data*, pages 25–71, 2006. (Cited on page 28.)

[39] H. Berlioz. *Traité d'instrumentation et d'orchestration*. Gregg International Publishers, 1855. (Cited on pages 3, 6, and 9.)

[40] D. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In *AAAI-94 workshop on knowledge discovery in databases*, pages 229–248, 1994. (Cited on page 41.)

[41] T. Bernecker, M. Houle, H.P. Kriegel, P. Kroger, M. Renz, E. Schubert, and A. Zimek. Quality of similarity rankings in time series. *Advances in Spatial and Temporal Databases*, pages 422–440, 2011. (Cited on pages 93 and 192.)

[42] S. Berretti, A. Del Bimbo, and P. Pala. Retrieval by shape similarity with perceptual distance and effective indexing. *IEEE Transactions on multimedia*, 2(4): 225–239, 2000. (Cited on page 49.)

[43] R. Bhargava, H. Kargupta, and M. Powers. Energy consumption in data analysis for on-board and distributed applications. In *Proceedings of the ICML*, volume 3, 2003. (Cited on page 48.)

[44] V. Bhaskar, S.K. Gupta, and A.K. Ray. Applications of multiobjective optimization in chemical engineering. *Reviews in Chemical Engineering*, 16(1):1–54, 2000. ISSN 0167-8299. (Cited on pages 50 and 54.)

[45] M. Bicego, V. Murino, and M. Figueiredo. Similarity-based clustering of sequences using hidden Markov models. *Lecture Notes in Computer Science*, 2743:95–104, 2003. (Cited on page 44.)

[46] L. Biel, O. Pettersson, L. Philipson, and P. Wide. Ecg analysis: a new approach in human identification. *Instrumentation and Measurement, IEEE Transactions on*, 50 (3):808–812, 2001. (Cited on page 113.)

[47] V.L. Blankers, C. Heuvel, K.Y. Franke, and L.G. Vuurpijl. Icdar 2009 signature verification competition. In *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*, pages 1403–1407. IEEE, 2009. (Cited on page 129.)

[48] B. Blankertz, K.R. Muller, G. Curio, T.M. Vaughan, G. Schalk, J.R. Wolpaw, A. Schlogl, C. Neuper, G. Pfurtscheller, T. Hinterberger, et al. The bci competition 2003: progress and perspectives in detection and discrimination of eeg single trials. *Biomedical Engineering, IEEE Transactions on*, 51(6):1044–1051, 2004. (Cited on pages 93 and 189.)

[49] B. Blankertz, G. Dornhege, M. Krauledat, K.R. Muller, and G. Curio. The non-invasive berlin brain-computer interface: Fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37(2):539–550, 2007. (Cited on pages 93 and 189.)

[50] C. Bohm, S. Berchtold, and D.A. Keim. Searching in high-dimensional spaces: Index structures for improving the performance of multimedia databases. *ACM Computing Surveys*, 33(3):322–373, 2001. (Cited on page 46.)

[51] B. Bollobas, G. Das, D. Gunopulos, and H. Mannila. Time-series similarity problems and well-separated geometric sets. In *Proceedings of the 13th symposium on computational geometry*, pages 454–456, 1997. (Cited on page 42.)

[52] M. Botta, A. Giordana, and L. Saitta. Learning fuzzy concept definitions. In *Fuzzy Systems, 1993., Second IEEE International Conference on*, pages 18–22. IEEE, 1993. (Cited on pages 93 and 187.)

[53] R. Bousseljot, D. Kreiseler, and A. Schnabel. Nutzung der ekg-signaldatenbank cardiodat der ptb uber das internet. *Biomedizinische Technik/Biomedical Engineering*, 40(s1):317–318, 1995. (Cited on page 197.)

[54] K.W. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3d and multi-modal 3d+ 2d face recognition. *Computer Vision and Image Understanding*, 101(1):1–15, 2006. (Cited on page 129.)

[55] G.E.P. Box, G.M. Jenkins, and G.C. Reinsel. *Time series analysis: forecasting and control*. Holden-day San Francisco, 1976. (Cited on page 32.)

[56] A. Brancucci and P. San Martini. Laterality in the perception of temporal cues of musical timbre. *Neuropsychologia*, 37(13):1445–1451, 1999. (Cited on page 11.)

[57] AS Bregman. Timbre, orchestration, dissonance, et organisation auditive. *Le timbre, métaphore pour la composition*, pages 204–215, 1991. (Cited on pages 5 and 9.)

[58] P.J. Brockwell and R.A. Davis. *Introduction to time series and forecasting*. Springer Verlag, 2002. (Cited on page 31.)

[59] P.J. Brockwell and R.A. Davis. *Time series: theory and methods*. Springer Verlag, 2009. ISBN 1441903194. (Cited on page 31.)

[60] R.A. Brown and R. Frayne. A fast discrete s-transform for biomedical signal processing. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 2586–2589. IEEE, 2008. (Cited on pages 117 and 205.)

[61] M. Bsoul, H. Minn, M. Nourani, G. Gupta, and L. Tamil. Real-time sleep quality assessment using single-lead ecg and multi-stage svm classifier. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, volume 2010, page 1178, 2010. (Cited on pages 93 and 199.)

[62] J. Buhler and M. Tompa. Finding motifs using random projections. *Journal of computational biology*, 9(2):225–242, 2002. ISSN 1066-5277. (Cited on page 34.)

[63] H.S. Burkom, S.P. Murphy, and G. Shmueli. Automated time series forecasting for biosurveillance. *Statistics in Medicine*, 26(22):4202–4218, 2007. (Cited on pages 23 and 32.)

[64] R. Cai, L. Lu, and H.J. Zhang. Using structure patterns of temporal and spectral feature in audio similarity measure. In *Proceedings of the 11th ACM international conference on Multimedia*, pages 219–222, 2003. (Cited on pages 71 and 132.)

[65] Y. Cai and R. Ng. Indexing spatio-temporal trajectories with Chebyshev polynomials. In *Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pages 599–610, Paris, France, 2004. ACM. (Cited on page 36.)

[66] L.M. Camarinha-Matos, L.S. Lopes, and J. Barata. Integration and learning in supervision of flexible assembly systems. *Robotics and Automation, IEEE Transactions on*, 12(2):202–219, 1996. (Cited on pages 93 and 197.)

[67] P. Cano and M. Koppenberger. Automatic sound annotation. In *Proceedings of the 14th IEEE Workshop on Machine Learning for Signal Processing*, pages 391–400. IEEE, 2004. (Cited on page 139.)

[68] P. Cano, M. Koppenberger, S. Le Groux, J. Ricard, P. Herrera, and N. Wack. Nearest-neighbor generic sound classification with a wordnet-based taxonomy. In *Proceedings of the 116th AES Convention*, Berlin, Germany, 2004. Citeseer. (Cited on page 71.)

[69] L. Cao and F. Tay. Feature selection for support vector machines in financial time series forecasting. *Intelligent Data Engineering and Automated Learning. Lecture Notes in Computer Science*, 1983:41–65, 2009. (Cited on page 32.)

[70] R. Cappelli, M. Ferrara, and D. Maltoni. Fingerprint indexing based on minutia cylinder-code. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33 (5):1051–1057, 2011. (Cited on page 129.)

[71] G. Carpentier. *Approche computationnelle de l'orchestration musciale-Optimisation multicritère sous contraintes de combinaisons instrumentales dans de grandes banques de sons*. PhD thesis, Université Pierre et Marie Curie-Paris VI, 2008. (Cited on pages 150, 153, 158, and 159.)

[72] G. Carpentier, D. Tardieu, G. Assayag, X. Rodet, and E. Saint-James. Imitative and generative orchestrations using pre-analyzed sound databases. In *Proceedings of Sound and Music Computing Conference (SMC)(Marseille, France)*, pages 115–122, 2006. (Cited on page 150.)

[73] A. Casella, V. Mortari, and P. Petit. *La Technique de l'orchestre contemporain*. Ricordi, 1958. (Cited on pages 3 and 4.)

[74] M. Casey. General sound classification and similarity in mpeg-7. *Organised Sound*, 6(02):153–164, 2001. (Cited on page 132.)

[75] M.A. Casey. Reduced-rank spectra and minimum-entropy priors as consistent and reliable cues for generalized sound recognition. In *Workshop for Consistent & Reliable Acoustic Cues*, 2001. (Cited on page 132.)

[76] M.A. Casey. Acoustic lexemes for organizing internet audio. *Contemporary Music Review*, 24(6):489–508, 2005. (Cited on page 72.)

[77] M.A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. Content-based music information retrieval: current directions and future challenges. *Proceedings of the IEEE*, 96(4):668–696, 2008. (Cited on page 71.)

[78] M.J. Castro-Bleda, S. Espana, J. Gorbe, F. Zamora, D. Llorens, A. Marzal, F. Prat, and J.M. Vilar. Improving a dtw-based recognition engine for on-line handwritten characters by using mlps. In *10th International Conference on Document Analysis and Recognition*, pages 1260–1264, 2009. (Cited on pages 93 and 195.)

[79] K. Chakrabarti and S. Mehrotra. The hybrid tree: an index structure for high dimensional feature spaces. In *Data Engineering, 1999. Proceedings., 15th International Conference on*, pages 440–447, 1999. (Cited on page 46.)

[80] F.K.P. Chan, A.W.C. Fu, and C. Yu. Haar wavelets for efficient similarity search of time-series: with and without time warping. *IEEE Transactions on knowledge and data engineering*, 15(3):686–705, 2003. (Cited on page 36.)

[81] K Chan and AW Fu. Efficient time series matching by wavelets. In *Proceedings of the 15th IEEE International conference on data engineering*, pages 126 – 133, Sydney, Australia, 1999. (Cited on pages 26 and 36.)

[82] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):15, 2009. (Cited on page 33.)

[83] V. Chankong and Y.Y. Haimes. *Multiobjective decision making: theory and methodology*. North-Holland New York, 1983. ISBN 0444007105. (Cited on page 50.)

[84] JC Chappelier and A. Grumbach. A Kohonen map for temporal sequences. In *Proceedings of the Conference on Neural Networks and Their Applications.*, pages 104–110, 1996. (Cited on page 28.)

[85] L. Chen and R. Ng. On the marriage of Lp-norms and edit distance. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, pages 792–803. VLDB Endowment, 2004. (Cited on page 42.)

[86] L. Chen, M.T. Ozsu, and V. Oria. Robust and fast similarity search for moving object trajectories. In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pages 491–502, Baltimore, Maryland, USA, 2005. ACM. (Cited on page 42.)

[87] Q. Chen, L. Chen, X. Lian, Y. Liu, and J.X. Yu. Indexable PLA for efficient similarity search. In *Proceedings of the 33rd international conference on Very large data bases*, pages 435–446. VLDB Endowment, 2007. (Cited on page 38.)

[88] X. Chen and Y. Zhan. Multi-scale anomaly detection algorithm based on infrequent pattern of time series. *Journal of Computational and Applied Mathematics*, 214 (1):227–237, 2008. (Cited on page 33.)

[89] X. Chen, S. Kwong, and M. Li. A compression algorithm for DNA sequences and its applications in genome comparison. In *Proceedings of the fourth annual international conference on Computational molecular biology*, page 107, 2000. (Cited on page 44.)

[90] Y. Chen, MA Nascimento, B.C. Ooi, and AKH Tung. Spade: On shape-based pattern detection in streaming time series. In *IEEE 23rd International Conference on Data Engineering, 2007*, pages 786–795, 2007. (Cited on page 42.)

[91] V. Chhieng and R. Wong. Adaptive distance measurement for time series databases. *Lecture Notes in Computer Science*, 4443:598–610, 2010. (Cited on page 43.)

[92] B. Chiu, E. Keogh, and S. Lonardi. Probabilistic discovery of time series motifs. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 493–498, Washington, D.C, USA, 2003. ACM. (Cited on page 34.)

[93] M. Chuah and F. Fu. ECG anomaly detection via time series analysis. In *Frontiers of High Performance Computing and Networking ISPA 2007 Workshops*, pages 123–135. Springer, 2007. (Cited on pages 33 and 47.)

[94] W.J. Conover. Practical nonparametric statistics, 1980. (Cited on pages 94 and 135.)

[95] D. Coomans and DL Massart. Alternative< i> k</i>-nearest neighbour rules in supervised pattern recognition: Part 1. k-nearest neighbour classification by using alternative voting rules. *Analytica Chimica Acta*, 136:15–27, 1982. (Cited on page 89.)

[96] M. Corduas and D. Piccolo. Time series clustering and classification by the autoregressive metric. *Computational statistics & data analysis*, 52(4):1860–1872, 2008. (Cited on page 28.)

[97] G. Cormode, S. Muthukrishnan, and W. Zhuang. Conquering the divide: Continuous clustering of distributed data streams. In *IEEE 23rd International Conference on Data Engineering, 2007*, pages 1036–1045, 2007. (Cited on page 28.)

[98] C. Costa Santos, J. Bernardes, PMB Vitanyi, and L. Antunes. Clustering fetal heart rate tracings by compression. In *19th International Symposium on Computer-Based Medical Systems*, pages 685–690, 2006. (Cited on page 44.)

[99] G. Das, D. Gunopulos, and H. Mannila. Finding similar time series. In *Principles of data mining and knowledge discovery: First European Symposium, PKDD'97, June 24-27*, volume 1263, pages 88–100, Trondheim, Norway, 1997. Springer Verlag. (Cited on page 42.)

[100] S. Davis, A. Jacobson, D. Suszcynsky, M. Cai, and D. Eads. Forte time series dataset, 2005. URL `http://nis-www.lanl.gov/~eads/datasets/forte`. (Cited on page 192.)

[101] M. Degli Esposti, C. Farinelli, and G. Menconi. Sequence distance via parsing complexity: Heartbeat signals. *Chaos, Solitons & Fractals*, 39(3):991–999, 2009. (Cited on page 44.)

[102] J. Demsar. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*, 7:1–30, 2006. (Cited on pages 92, 94, 96, 98, and 175.)

[103] K. Deng, AW Moore, and MC Nechyba. Learning to recognize time series: combining ARMA models with memory-based learning. In *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation, 1997. CIRA'97*, pages 246–251, 1997. (Cited on page 30.)

[104] A. Denton. Kernel-density-based clustering of time series subsequences using a continuous random-walk noise model. In *Proceedings of the fifth IEEE International Conference on Data Mining*, pages 122–129, 2005. (Cited on page 29.)

[105] D.B. Dias, R.C.B. Madeo, T. Rocha, H.H. Biscaro, and S.M. Peres. Hand movement recognition for brazilian sign language: A study using distance-based neural networks. In *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, pages 697–704. IEEE, 2009. (Cited on page 194.)

[106] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh. Querying and mining of time series data: experimental comparison of representations and distance measures. *Proceedings of the VLDB Endowment*, 1(2):1542–1552, 2008. (Cited on pages 36, 39, 41, 43, 44, 49, and 89.)

[107] P. Domingos and G. Hulten. Mining high-speed data streams. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 71–80. ACM, 2000. (Cited on page 47.)

[108] G. Dong, J. Han, L.V.S. Lakshmanan, J. Pei, H. Wang, and P.S. Yu. Online mining of changes from data streams: Research problems and preliminary results. In *Proceedings of the 2003 ACM SIGMOD Workshop on Management and Processing of Data Streams*, San Diego, CA, 2003. (Cited on page 49.)

[109] B. Dorizzi, R. Cappelli, M. Ferrara, D. Maio, D. Maltoni, N. Houmani, S. Garcia-Salicetti, and A. Mayoue. Fingerprint and on-line signature verification competitions at icb 2009. *Advances in Biometrics*, pages 725–732, 2009. (Cited on page 129.)

[110] J.S. Downie. The scientific evaluation of music information retrieval systems: Foundations and future. *Computer Music Journal*, 28(2):12–23, 2004. (Cited on pages 15, 57, 72, and 94.)

[111] J.S. Downie and S.J. Cunningham. Toward a theory of music information retrieval queries: System design implications. In *Proceedings: Third International Conference on Music Information Retrieval. ISMIR*, volume 2002, pages 13–17, 2002. (Cited on page 57.)

[112] H.W. Draper, C.J. Peffer, F.W. Stallmann, D. Littmann, and H.V. Pipberger. The corrected orthogonal electrocardiogram and vectorcardiogram in 510 normal men (frank lead system). *Circulation*, 30(6):853–864, 1964. (Cited on page 131.)

[113] F.Y. Edgeworth. Mathematical psychics. *History of Economic Thought Books*, 1881. (Cited on page 50.)

[114] J. Eggermont, J.N. Kok, and W.A. Kosters. Genetic programming for data classification: Partitioning the search space. In *Proceedings of the 2004 ACM symposium on Applied computing*, pages 1001–1005. ACM, 2004. (Cited on pages 93 and 194.)

[115] M. Ehrgott. *Multicriteria optimization*, volume 491. Springer Verlag, 2005. (Cited on page 50.)

[116] M. Ehrgott and X. Gandibleux. A survey and annotated bibliography of multiobjective combinatorial optimization. *Or Spectrum*, 22(4):425–460, 2000. (Cited on page 54.)

[117] N. El-Bendary, H. Al-Qaheri, HM Zawbaa, M. Hamed, A.E. Hassanien, Q. Zhao, and A. Abraham. Hsas: Heart sound authentication system. In *Nature and Biologically Inspired Computing (NaBIC), 2010 Second World Congress on*, pages 351–356. IEEE, 2010. (Cited on pages 113 and 204.)

[118] G. Ellis and A. Dix. An explorative analysis of user evaluation studies in information visualisation. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pages 1–7. ACM, 2006. (Cited on page 79.)

[119] C. Faloutsos and V. Megalooikonomou. On data mining, compression, and kolmogorov complexity. *Data Mining and Knowledge Discovery*, 15(1):3–20, 2007. (Cited on page 48.)

[120] C Faloutsos, M Ranganathan, and Y Manolopulos. Fast subsequence matching in time-series databases. *SIGMOD Record*, 23:419 – 429, 1994. (Cited on pages 23, 36, 44, 46, and 47.)

[121] S.Z. Fatemian, F. Agrafioti, and D. Hatzinakos. Heartid: Cardiac biometric recognition. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1–5. IEEE, 2010. (Cited on page 204.)

[122] P. Ferreira, P. Azevedo, C. Silva, and R. Brito. Mining approximate motifs in time series. In *Lecture Notes in Computer Science*, volume 4265, pages 89–101. Springer, 2006. (Cited on page 34.)

[123] J. Figueira, S. Greco, and M. Ehrgott. *Multiple criteria decision analysis: state of the art surveys*. Springer Verlag, 2005. ISBN 038723067X. (Cited on page 50.)

[124] J.A. Flanagan. A non-parametric approach to unsupervised learning and clustering of symbol strings and sequences. In *Proceedings of the 4th Workshop on Self-Organizing Maps (WSOM03)*, pages 128–133, 2003. (Cited on page 43.)

[125] C.M. Fonseca, L. Paquete, and M. López-Ibánez. An improved dimension-sweep algorithm for the hypervolume indicator. In *IEEE Congress on Evolutionary Computation (CEC'2006)*, pages 1157–1163, 2006. (Cited on page 84.)

[126] A. Frank and A. Asuncion. UCI machine learning repository, 2010. URL `http://archive.ics.uci.edu/ml`. (Cited on pages 93, 187, 188, 191, 193, 194, 195, 196, 197, 199, 200, and 201.)

[127] A.L. Freire, G.A. Barreto, M. Veloso, and A.T. Varela. Short-term memory mechanisms in neural network learning of robot navigation tasks: a case study. In *Proceedings of the 6th Latin American Robotics Symposium (LARS2009)*, 2009. (Cited on pages 93 and 201.)

[128] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, and J.M. Hausdorff. Treadmill walking as an external pacemaker to improve gait rhythm and stability in parkinson's disease. *Movement disorders*, 20(9):1109–1114, 2005. (Cited on page 193.)

[129] E. Frentzos, K. Gratsias, and Y. Theodoridis. Index-based most similar trajectory search. In *IEEE 23rd International Conference on Data Engineering, 2007. ICDE 2007*, pages 816–825, 2007. (Cited on page 42.)

[130] S. Fröhwirth-Schnatter and S. Kaufmann. Model-based clustering of multiple time series. *Journal of Business and Economic Statistics*, 26(1):78–89, 2008. (Cited on page 28.)

[131] A.W. Fu, E. Keogh, L.Y. Lau, C.A. Ratanamahatana, and R.C.W. Wong. Scaling and time warping in time series querying. *The VLDB Journal - The International Journal on Very Large Data Bases*, 17(4):921, 2008. (Cited on page 42.)

[132] E. Fuchs, T. Gruber, H. Pree, and B. Sick. Temporal data mining using shape space representations of time series. *Neurocomputing*, 74(1-3):379–393, 2010. (Cited on page 38.)

[133] M.M. Gaber, A. Zaslavsky, and S. Krishnaswamy. Mining data streams: a review. *ACM Sigmod Record*, 34(2):18–26, 2005. (Cited on page 47.)

[134] S. Gaffney and P. Smyth. Trajectory clustering with mixtures of regression models. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 63–72. ACM, 1999. (Cited on pages 28 and 44.)

[135] X. Ge and P. Smyth. Deformable Markov model templates for time-series pattern matching. In *Proceedings of the 6th ACM International conference on Knowledge Discovery and Data Mining*, pages 81–90, 2000. (Cited on page 43.)

[136] P. Geurts. Pattern extraction for time series classification. In *Proceedings of the 5th European conference on principles of data mining and knowledge discovery*, pages 115 – 127, Freiburg, Germany, 2001. (Cited on page 30.)

[137] A. Ghias, J. Logan, D. Chamberlin, and B.C. Smith. Query by humming: musical information retrieval in an audio database. In *Proceedings of the third ACM international conference on Multimedia*, pages 231–236. ACM, 1995. (Cited on pages 71 and 132.)

[138] R. Giot, M. El-Abed, and C. Rosenberger. Greyc keystroke: a benchmark for keystroke dynamics biometric systems. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*, pages 1–6. IEEE, 2009. (Cited on page 129.)

[139] L. Golab and M.T. Ozsu. Issues in data stream management. *ACM Sigmod Record*, 32(2):5–14, 2003. (Cited on page 47.)

[140] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.K. Peng, and H.E. Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000. (Cited on pages 190, 193, 197, 199, and 200.)

[141] D. Goldin and P. Kanellakis. On similarity queries for time-series data: Constraint specification and implementation. In *Principles and Practice of Constraint Programming - CP95*, pages 137–153. Springer, 1995. (Cited on pages 27, 35, and 41.)

[142] D.Q. Goldin, T.D. Millstein, and A. Kutlu. Bounded similarity querying for time-series data. *Information and Computation*, 194(2):203–241, 2004. (Cited on page 35.)

[143] JA Gómez-Limón and J. Berbel. Multicriteria analysis of derived water demand functions: a Spanish case study. *Agricultural Systems*, 63(1):49–72, 2000. ISSN 0308-521X. (Cited on page 54.)

[144] J.W. Gordon. The perceptual attack time of musical tones. *The Journal of the Acoustical Society of America*, 82:88, 1987. (Cited on page 10.)

[145] D.O. Gorodnichy. Evolution and evaluation of biometric systems. In *Computational Intelligence for Security and Defense Applications, 2009. CISDA 2009. IEEE Symposium on*, pages 1–8. IEEE, 2009. (Cited on pages 121 and 122.)

[146] M Goto, H Hashiguchi, T Nishimura, and R Oka. Rwc music database : music genre database and musical instrument sound database. In *Proceedings of the 4th International Conference on Music Information Retrieval*, pages 229 – 230, Washington, USA, 2003. (Cited on page 68.)

[147] S.D. Greenwald, P. Albrecht, G.B. Moody, and R.G. Mark. Estimating confidence limits for arrhythmia detector performance. *Computers in Cardiology*, 12:383–386, 1985. (Cited on page 200.)

[148] J.M. Grey. Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61(5):1270–1277, 1977. (Cited on page 15.)

[149] S. Gudmundsson, T.P. Runarsson, and S. Sigurdsson. Support vector machines and dynamic time warping for time series. In *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, pages 2772–2776. IEEE, 2008. (Cited on pages 83, 89, 95, and 96.)

[150] S. Gudmundsson, T.P. Runarsson, and S. Sigurdsson. Test-retest reliability and feature selection in physiological time series classification. *Computer Methods and Programs in Biomedicine*, 105(1):50–60, 2012. (Cited on pages 83, 89, 93, and 197.)

[151] C. Guger, W. Domej, G. Lindner, K. Pfurtscheller, G. Pfurtscheller, and G. Edlinger. Effects of a fast cable car ascent to an altitude of 2700 meters on eeg and ecg. *Neuroscience letters*, 377(1):53–58, 2005. (Cited on pages 93 and 191.)

[152] F. Gullo, G. Ponti, A. Tagarelli, and S. Greco. A time series representation model for accurate and fast similarity detection. *Pattern Recognition*, 42(11):2998–3014, 2009. (Cited on page 38.)

[153] G. Guo and S.Z. Li. Content-based audio classification and retrieval by support vector machines. *Neural Networks, IEEE Transactions on*, 14(1):209–215, 2003. (Cited on pages 71, 133, 138, and 139.)

[154] G. Guo, H.J. Zhang, and S.Z. Li. Boosting for content-based audio classification and retrieval: an evaluation. In *IEEE International Conference on Multimedia and Expo (ICME 2001)*, pages 997–1000, 2001. (Cited on pages 132, 133, 138, and 139.)

[155] S. Gupta, A. Ray, and E. Keller. Symbolic time series analysis of ultrasonic data for early detection of fatigue damage. *Mechanical Systems and Signal Processing*, 21(2):866–884, 2007. (Cited on page 33.)

[156] D. Gusfield. *Algorithms on strings, trees, and sequences: computer science and computational biology*. Cambridge Univ Pr, 1997. (Cited on page 46.)

[157] N. Hammami and M. Bedda. Improved tree model for arabic speech recognition. In *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*, volume 5, pages 521–526. IEEE, 2010. (Cited on page 187.)

[158] N. Hammami and M. Sellam. Tree distribution classifier for automatic spoken arabic digit recognition. In *Internet Technology and Secured Transactions, 2009. ICITST 2009. International Conference for*, pages 1–4. IEEE, 2009. (Cited on pages 93 and 187.)

[159] J. Han and M. Kamber. *Data mining: concepts and techniques*. Morgan Kaufmann, 2006. ISBN 1558609016. (Cited on page 28.)

[160] R.I.D. Harris and R. Sollis. *Applied time series modelling and forecasting*. J. Wiley, 2003. ISBN 0470844434. (Cited on page 31.)

[161] K. Hayashi. Multicriteria analysis for agricultural resource management : a critical survey and future perspectives. *European Journal of Operational Research*, 122(2):486–500, 2000. ISSN 0377-2217. (Cited on page 54.)

[162] G. Hebrail and B. Hugueney. Symbolic representation of long time-series. In *Symbolic Data Analysis at the 4th European Conference on Principles of Data Mining and Knowledge Discovery*, pages 56–65, 2000. (Cited on page 28.)

[163] M. Helen and T. Lahti. Query by example methods for audio signals. In *Proceedings of the 7th Nordic Signal Processing Symposium*, pages 302–305, 2006. (Cited on pages 133 and 139.)

[164] M. Helen and T. Lahti. Query by example in large databases using key-sample distance transformation and clustering. In *Proceedings of the 9th IEEE International Symposium on Multimedia Workshops (ISMW'07)*, pages 303–308, 2007. (Cited on page 72.)

[165] M. Helen and T. Virtanen. Query by example of audio signals using euclidean distance between gaussian mixture models. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages 225–228, 2007. (Cited on pages 71 and 132.)

[166] J.M. Hellerstein, E. Koutsoupias, and C.H. Papadimitriou. On the analysis of indexing schemes. In *Proceedings of the 16th ACM Symposium on Principles of Database Systems*, pages 249–256, 1997. (Cited on page 47.)

[167] H.L.F. Helmholtz and A.J. Ellis. The sensations of tone: As a physiological basis for the theory of music. 1875. (Cited on page 9.)

[168] LJ Herrera, H. Pomares, I. Rojas, A. Guillén, A. Prieto, and O. Valenzuela. Recursive prediction for long term time series forecasting using advanced models. *Neurocomputing*, 70(16-18):2870–2880, 2007. (Cited on page 32.)

[169] J. Himberg, K. Korpiaho, J. Tikanmaki, and H.T.T. Toivonen. Time series segmentation for context recognition in mobile devices. In *Proceedings of the 1st IEEE International Conference on Data Mining*, pages 203–210, 2001. (Cited on page 31.)

[170] J. Himberg, J. Mantyjarvi, and P. Korpipaa. Using PCA and ICA for exploratory data analysis in situation awareness. In *Proceedings of the International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 127–131, 2001. (Cited on page 35.)

[171] C. Holland and O.V. Komogortsev. Biometric identification via eye movement scanpaths in reading. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–8. IEEE, 2011. (Cited on page 129.)

[172] Y.W. Huang and P.S. Yu. Adaptive query processing for time-series data. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 282–286. ACM, 1999. (Cited on page 38.)

[173] G. Hulten, L. Spencer, and P. Domingos. Mining time-changing data streams. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 97–106. ACM, 2001. (Cited on page 47.)

[174] T.A. Hummel. Simulation of human voice timbre by orchestration of acoustic music instruments. In *Proceedings of International Computer Music Conference (ICMC)*, 2005. (Cited on page 150.)

[175] Y. Ichimaru and GB Moody. Development of the polysomnographic database on cd-rom. *Psychiatry and Clinical Neurosciences*, 53(2):175–177, 1999. (Cited on page 199.)

[176] P. Indyk, N. Koudas, and S. Muthukrishnan. Identifying representative trends in massive time series data sets using sketches. In *Proceedings of the 26th International Conference on Very Large Data Bases*, pages 363–372. Morgan Kaufmann Publishers Inc., 2000. (Cited on pages 26 and 36.)

[177] K.T.U. Islam, K. Hasan, Y.K. Lee, and S. Lee. Enhanced 1-nn time series classification using badness of records. In *Proceedings of the 2nd international conference on Ubiquitous information management and communication*, pages 108–113. ACM, 2008. (Cited on pages 83 and 89.)

[178] S.A. Israel, J.M. Irvine, A. Cheng, M.D. Wiederhold, and B.K. Wiederhold. Ecg to identify individuals. *Pattern Recognition*, 38(1):133–142, 2005. (Cited on page 113.)

[179] P. Iverson and C.L. Krumhansl. Isolating the dynamic attributes of musical timbre. *The Journal of the Acoustical Society of America*, 94:2595, 1993. (Cited on page 10.)

[180] A. Jain, L. Hong, and S. Pankanti. Biometric identification. *Communications of the ACM*, 43(2):90–98, 2000. (Cited on pages 112, 113, and 131.)

[181] A.K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, 2004. (Cited on page 113.)

[182] GJ Janacek, AJ Bagnall, and M. Powell. A likelihood ratio distance measure for the similarity between the Fourier transform of time series. *Lecture Notes in Computer Science*, 3518:737–743, 2005. (Cited on page 43.)

[183] J. Jasper and K.R. Othman. Feature extraction for human identification based on envelogram signal analysis of cardiac sounds in time-frequency domain. In *Electronics and Information Engineering (ICEIE), 2010 International Conference On*, volume 2, pages V2–228. IEEE, 2010. (Cited on pages 113 and 204.)

[184] C. Jeffery. Synthetic lightning emp data, 2005. URL `http://nis-www.lanl.gov/~eads/datasets/emp`. (Cited on pages 93 and 199.)

[185] S.L. Jeng and Y.T. Huang. Time Series Classification Based on Spectral Analysis. *Communications in Statistics-Simulation and Computation*, 37(1):132–142, 2008. (Cited on page 30.)

[186] M.W. Kadous. Grasp: Recognition of australian sign language using instrumented gloves, 1995. (Cited on pages 93 and 188.)

[187] M.W. Kadous. *Temporal Classification: Extending the Classification Paradigm to Multivariate Time Series*. PhD thesis, The University of New South Wales, 2002. (Cited on pages 93 and 188.)

[188] K. Kalpakis, D. Gada, and V. Puttagunta. Distance measures for effective clustering of ARIMA time-series. In *Proceedings of the IEEE International Conference on Data Mining*, pages 273–280, 2001. (Cited on page 39.)

[189] B. Kaluza, V. Mirchevska, E. Dovgan, M. Lustrek, and M. Gams. An agent-based approach to care in independent living. *Ambient Intelligence*, pages 177–186, 2010. (Cited on pages 93 and 196.)

[190] S. Kanade, D. Petrovska-Delacrétaz, and B. Dorizzi. Cancelable iris biometrics and using error correcting codes to reduce variability in biometric data. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 120–127. IEEE, 2009. (Cited on page 129.)

[191] H. Kauppinen, T. Seppanen, and M. Pietikainen. An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):201–207, 1995. (Cited on page 49.)

[192] A. Kehagias. A hidden Markov model segmentation procedure for hydrological and environmental time series. *Stochastic Environmental Research and Risk Assessment*, 18(2):117–130, 2004. (Cited on page 31.)

[193] E Keogh and S Kasetty. On the need for time series data mining benchmarks : a survey and empirical demonstration. In *Proceedings of the 8th ACM SIGKDD International conference on knowledge discovery and data mining*, pages 102 – 111, Edmonton, Alberta, Canada, 2002. (Cited on pages 49 and 92.)

[194] E. Keogh and S. Kasetty. On the need for time series data mining benchmarks: A survey and empirical demonstration. *Data Mining and Knowledge Discovery*, 7(4): 349–371, 2003. (Cited on pages 35, 39, 40, 44, 46, 94, and 95.)

[195] E. Keogh and M. Pazzani. An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. In *Proceedings of the 4th International Conference of Knowledge Discovery and Data Mining*, pages 239–241. AAAI Press, 1998. (Cited on pages 29, 31, 38, 41, and 48.)

[196] E. Keogh and C.A. Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7(3):358–386, 2005. (Cited on pages 27, 41, and 47.)

[197] E Keogh, K Chakrabarti, and M Pazzani. Locally adaptive dimensionality reduction for indexing large time series databases. In *Proceedings of ACM conference on management of data*, pages 151 – 162, 2001. (Cited on pages 26, 35, and 38.)

[198] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra. Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and Information Systems*, 3(3):263–286, 2001. (Cited on pages 26, 36, and 46.)

[199] E. Keogh, S. Chu, D. Hart, and M. Pazzani. Segmenting time series: A survey and novel approach. *Data mining in time series databases*, pages 1–21, 2003. (Cited on pages 23, 30, and 47.)

[200] E. Keogh, J. Lin, and W. Truppel. Clustering of time series subsequences is meaningless: implications for previous and future research. In *3rd IEEE International Conference on Data Mining*, pages 115–122, 2003. (Cited on pages 29 and 34.)

[201] E. Keogh, S. Lonardi, and C.A. Ratanamahatana. Towards parameter-free data mining. In *Proceedings of 10th ACM international conference on Knowledge discovery and data mining*, pages 206–215, 2004. (Cited on pages 26, 36, 44, and 48.)

[202] E. Keogh, J. Lin, S.H. Lee, and H.V. Herle. Finding the most unusual time series subsequence: algorithms and applications. *Knowledge and Information Systems*, 11 (1):1–27, 2007. (Cited on page 33.)

[203] G. Kerr, H.J. Ruskin, M. Crane, and P. Doolan. Techniques for clustering gene expression data. *Computers in Biology and Medicine*, 38(3):283–293, 2008. (Cited on page 29.)

[204] G.A. Kiker, T.S. Bridges, A. Varghese, T.P. Seager, and I. Linkov. Application of multicriteria decision analysis in environmental decision making. *Integrated Environmental Assessment and Management*, 1(2):95–108, 2005. ISSN 1551-3793. (Cited on page 54.)

[205] S.W. Kim, S. Park, and W.W. Chu. An Index-Based Approach for Similarity Search Supporting Time Warping in Large Sequence Databases. In *Proceedings of the 17th International Conference on Data Engineering*, pages 607–614. IEEE Computer Society, 2001. (Cited on page 27.)

[206] C. Kœchlin. *Traité de l'orchestration en quatre volumes*, volume 1. Éditions Max Eschig, 1954. (Cited on pages 3 and 4.)

[207] E. Koliopoulou. Etude de la description morphologique des sons environnementaux. Technical report, IRCAM, Paris, 2012. (Cited on pages 139 and 143.)

[208] M. Kontaki, A.N. Papadopoulos, and Y. Manolopoulos. Adaptive similarity search in streaming time series with sliding windows. *Data & Knowledge Engineering*, 63 (2):478–502, 2007. (Cited on pages 27 and 47.)

[209] M. Kontaki, A. Papadopoulos, and Y. Manolopoulos. Similarity Search in Time Series. *Handbook of Research on Innovations in Database Technologies and Applications*, pages 288–299, 2009. (Cited on page 27.)

[210] F. Korn, HV Jagadish, and C. Faloutsos. Efficiently supporting ad hoc queries in large datasets of time sequences. In *Proceedings of the 1997 ACM SIGMOD international conference on Management of data*, pages 289–300. ACM, 1997. (Cited on pages 26 and 36.)

[211] T. Koskela. *Neural network methods in analysing and modelling time varying processes*. PhD thesis, Helsinki University of Technology Laboratory of Computational Engineering, 2003. (Cited on page 32.)

[212] V. Krasteva and I. Jekova. Assessment of ecg frequency and morphology parameters for automatic classification of life-threatening cardiac arrhythmias. *Physiological measurement*, 26:707, 2005. (Cited on pages 93 and 200.)

[213] M. Kudo, J. Toyama, and M. Shimbo. Multidimensional curve classification using passing-through regions. *Pattern Recognition Letters*, 20(11-13):1103–1111, 1999. (Cited on pages 93 and 194.)

[214] N. Kumar, N. Lolla, E. Keogh, S. Lonardi, C. Ratanamahatana, and L. Wei. Time-series bitmaps: a practical visualization tool for working with large time series databases. In *SIAM 2005 Data Mining Conference*, pages 531–535, 2005. (Cited on page 39.)

[215] R.T.H. Laennec. *De l'auscultation médiate ou Traité du diagnostic des maladies des poumons et du coeur, fondé principalement sur ce nouveau moyen d'exploration*, volume 2. Brosson et Chaude, 1819. (Cited on page 202.)

[216] LJ Latecki, V. Megalooikonomou, Q. Wang, R. Lakaemper, CA Ratanamahatana, and E. Keogh. Elastic partial matching of time series. *Knowledge Discovery in Databases*, pages 577–584, 2005. (Cited on page 42.)

[217] L.J. Latecki, Q. Wang, S. Koknar-Tezel, and V. Megalooikonomou. Optimal subsequence bijection. In *IEEE Int. Conf. on Data Mining (ICDM)*, pages 565–570, Omaha, USA, 2007. (Cited on page 42.)

[218] MH Law and JT Kwok. Rival penalized competitive learning for model-based sequence clustering. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 2186–2195, 2000. (Cited on page 44.)

[219] S.H. Lee and J.S. Lim. Parkinson's disease classification using gait characteristics and wavelet-based feature extraction. *Expert Systems with Applications: An International Journal*, 39(8):7338–7344, 2012. (Cited on pages 93 and 193.)

[220] M. Lesaffre, M. Leman, K. Tanghe, B. De Baets, H. De Meyer, and J.P. Martens. User-dependent taxonomy of musical features as a conceptual framework for musical audio-mining technology. In *Proc. of the Stockholm Music Acoustics Conference*, 2003. (Cited on page 72.)

[221] C.S. Li, P.S. Yu, and V. Castelli. MALM: a framework for mining sequence database at multiple abstraction levels. In *Proceedings of the seventh international conference on Information and knowledge management*, pages 267–272. ACM, 1998. (Cited on page 31.)

[222] G. Li and A.A. Khokhar. Content-based indexing and retrieval of audio data using wavelets. In *IEEE International Conference on Multimedia and Expo*, volume 2, pages 885–888. IEEE, 2000. (Cited on pages 71 and 132.)

[223] S.Z. Li. Content-based audio classification and retrieval using the nearest feature line method. *Speech and Audio Processing, IEEE Transactions on*, 8(5):619–625, 2000. (Cited on pages 71, 132, 133, 138, and 139.)

[224] Y. Li and Y. Hou. Search audio data with the wavelet pyramidal algorithm. *Information processing letters*, 91(1):49–55, 2004. (Cited on page 132.)

[225] X. Lian and L. Chen. Efficient similarity search over future stream time series. *IEEE Transactions on Knowledge and Data Engineering*, 20(1):40–54, 2007. (Cited on pages 27 and 48.)

[226] X. Lian, L. Chen, and B. Wang. Approximate similarity search over multiple stream time series. *Lecture Notes in Computer Science*, 4443:962–968, 2010. (Cited on page 27.)

[227] T.W. Liao. Clustering of time series data–a survey. *Pattern Recognition*, 38(11): 1857–1874, 2005. (Cited on page 28.)

[228] AWC Liew, SH Leung, and WH Lau. Fuzzy image clustering incorporating spatial continuity. *IEEE Proceedings on Vision, Image and Signal Processing*, 147(2): 185–192, 2000. (Cited on page 49.)

[229] J. Lin and E. Keogh. Clustering of time-series subsequences is meaningless: implications for previous and future research. *Knowledge and information systems*, 8(2):154–177, 2005. (Cited on pages 23, 35, and 36.)

[230] J. Lin and Y. Li. Finding structural similarity in time series data using bag-of-patterns representation. In *Scientific and Statistical Database Management: 21st International Conference, SSDBM 2009 New Orleans, La, USA, June 2-4, 2009 Proceedings*, pages 461–477. Springer, 2009. (Cited on page 43.)

[231] J. Lin, E. Keogh, S. Lonardi, and B. Chiu. A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, pages 2–11. ACM New York, NY, USA, 2003. (Cited on pages 27, 38, and 73.)

[232] J. Lin, E. Keogh, S. Lonardi, J.P. Lankford, and D.M. Nystrom. Visually mining and monitoring massive time series. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 460–469. ACM, 2004. (Cited on page 23.)

[233] J. Lin, E. Keogh, L. Wei, and S. Lonardi. Experiencing sax: a novel symbolic representation of time series. *Data mining and knowledge discovery*, 15(2):107–144, 2007. (Cited on page 62.)

[234] T. Lin, N. Kaminski, and Z. Bar-Joseph. Alignment and classification of time series gene expression in clinical studies. *Bioinformatics*, 24(13):147–155, 2008. (Cited on pages 23 and 30.)

[235] Z. Liu, J.X. Yu, X. Lin, H. Lu, and W. Wang. *Locating motifs in time-series data*, pages 343–353. Springer, 2005. (Cited on page 34.)

[236] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 4:1–13, 2007. (Cited on page 30.)

[237] T. Lowitz, M. Ebert, W. Meyer, and B. Hensel. Hidden Markov Models for Classification of Heart Rate Variability in RR Time Series. In *World Congress on Medical Physics and Biomedical Engineering*, pages 1980–1983, Munich, Germany, 2009. Springer. (Cited on page 30.)

[238] A.A. Luisada, D.M. MacCanon, S. Kumar, and L.P. Feigen. Changing views on the mechanism of the first and second heart sounds. *American Heart Journal*, 88 (4):503–514, 1974. (Cited on page 113.)

[239] J. Ma and S. Perkins. Online novelty detection on temporal sequences. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 613–618. ACM, 2003. (Cited on page 33.)

[240] H. Mannila and JK Seppnen. Recognizing similar situations from event sequences. In *First SIAM Conference on Data Mining*, pages 1–16, Chicago, IL, USA, 2001. (Cited on page 43.)

[241] A.J. Mansfield and J.L. Wayman. Best practices in testing and reporting performance of biometric devices. *NPL Report CMSC*, 14(02), 2002. (Cited on page 120.)

[242] S. Mardle and S. Pascoe. A review of applications of multiple-criteria decision-making techniques to fisheries. *Marine Resource Economics*, 14:41–64, 1999. ISSN 0738-1360. (Cited on page 54.)

[243] P.F. Marteau. Time warp edit distance with stiffness adjustment for time series matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2): 306–318, 2008. (Cited on page 42.)

[244] S. McAdams and A. Bregman. Hearing musical streams. *Computer Music Journal*, 3(4):26–60, 1979. (Cited on page 15.)

[245] S. McAdams, S. Winsberg, S. Donnadieu, G. Soete, and J. Krimphoff. Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological research*, 58(3):177–192, 1995. (Cited on pages 15 and 72.)

[246] V. Megalooikonomou, G. Li, and Q. Wang. A dimensionality reduction technique for efficient similarity analysis of time series databases. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management*, pages 160–161, Washington, D.C., USA, 2004. ACM. (Cited on page 36.)

[247] V. Megalooikonomou, Q. Wang, G. Li, and C. Faloutsos. A multiresolution symbolic representation of time series. In *Proceedings. 21st International Conference on Data Engineering*, pages 668–679, 2005. (Cited on pages 36 and 43.)

[248] GA Mendoza and H. Martins. Multi-criteria decision analysis in natural resource management: a critical review of methods and new modelling paradigms. *Forest Ecology and Management*, 230(1-3):1–22, 2006. ISSN 0378-1127. (Cited on pages 50 and 54.)

[249] I. Mierswa and M. Wurst. Efficient case based feature construction. *Machine Learning: ECML 2005*, pages 641–648, 2005. (Cited on page 135.)

[250] J.R. Miller and E.C. Carterette. Perceptual space for musical structures. *The Journal of the Acoustical Society of America*, 58:711, 1975. (Cited on page 15.)

[251] PG Mills, RF Chamusco, S. Moos, and E. Craige. Echophonocardiographic studies of the contribution of the atrioventricular valves to the first heart sound. *Circulation*, 54(6):944–951, 1976. (Cited on page 113.)

[252] M. Misaki, Y. Kim, P.A. Bandettini, and N. Kriegeskorte. Comparison of multivariate classifiers and response normalizations for pattern-information fmri. *Neuroimage*, 53(1):103–118, 2010. (Cited on pages 92 and 94.)

[253] N. Misdariis, A. Minard, P. Susini, G. Lemaitre, S. McAdams, and E. Parizet. Environmental sound perception: Metadescription and modeling based on independent primary studies. *EURASIP Journal on Audio, Speech, and Music Processing*, 2010:26, 2010. (Cited on page 140.)

[254] B. Mjaaland, P. Bours, and D. Gligoroski. Nisk2009-gait mimicking-attack resistance testing of gait authentication systems. *Norsk informasjonssikkerhetskonferanse (NISK)*, 2009. (Cited on page 129.)

[255] Y. Mohammad and T. Nishida. Constrained Motif Discovery in Time Series. *New Generation Computing*, 27(4):319–346, 2009. (Cited on page 34.)

[256] F. Morchen, A. Ultsch, M. Thies, and I. Lohken. Modeling timbre distance with temporal statistics from polyphonic music. *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(1):81–90, 2006. (Cited on page 11.)

[257] M.D. Morse and J.M. Patel. An efficient and accurate method for evaluating time series similarity. In *Proceedings of the 2007 ACM international conference on Management of data*, pages 569–580, 2007. (Cited on page 42.)

[258] A. Mueen, E. Keogh, Q. Zhu, S. Cash, and B. Westover. Exact discovery of time series motifs. In *Proceedings of the SIAM International Conference on Data Mining (SDM)*, pages 473–484, 2009. (Cited on page 34.)

[259] M. Muhammad Fuad and P.F. Marteau. Extending the Edit Distance Using Frequencies of Common Characters. In *Proceedings of the 19th International Conference on Database and Expert Systems Applications*, pages 150–157, Turin, Italy, 2008. Springer. (Cited on page 42.)

[260] A. Nanopoulos, R. Alcock, and Y. Manolopoulos. Feature-based classification of time-series data. In *Information processing and technology*, pages 49–61, 2001. (Cited on pages 30 and 39.)

[261] Y. Ogras and H. Ferhatosmanoglu. Online summarization of dynamic time series data. *The VLDB Journal - The International Journal on Very Large Data Bases*, 15(1): 84–98, 2006. (Cited on page 31.)

[262] H.H. Otu and K. Sayood. A new sequence distance measure for phylogenetic tree construction. *Bioinformatics*, 19(16):2122–2130, 2003. (Cited on page 44.)

[263] R. Ouyang, L. Ren, W. Cheng, and C. Zhou. Similarity search and pattern discovery in hydrological time series data mining. *Hydrological Processes*, 24(9): 1198–1210, 2010. (Cited on page 23.)

[264] T. Palpanas, E. Keogh, V.B. Zordan, D. Gunopulos, and M. Cardle. Indexing large human-motion databases. In *Proceedings of the 13th international conference on Very large data bases*, pages 780–791, 2004. (Cited on page 35.)

[265] T. Palpanas, M. Vlachos, E. Keogh, D. Gunopulos, and W. Truppel. Online amnesic approximation of streaming time series. In *20th International Conference on data engineering*, pages 338–349, 2004. (Cited on page 38.)

[266] T. Palpanas, M. Vlachos, E. Keogh, and D. Gunopulos. Streaming time series summarization using user-defined amnesic functions. *IEEE Transactions on Knowledge and Data Engineering*, 20(7):992–1006, 2008. (Cited on page 31.)

[267] A. Panuccio, M. Bicego, and V. Murino. A Hidden Markov Model-based approach to sequential data clustering. *Lecture Notes in Computer Science*, 2396:734–743, 2002. (Cited on pages 39 and 43.)

[268] S. Papadimitriou and P. Yu. Optimal multi-scale patterns in time series streams. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pages 647–658, Chicago, IL, USA, 2006. (Cited on page 38.)

[269] S. Papadimitriou, J. Sun, and PS Yu. Local correlation tracking in time series. In *Data Mining, 2006. ICDM'06. Sixth International Conference on*, pages 456–465, 2006. (Cited on page 43.)

[270] V. Pareto. Cours d'Economie Politique, volume I and II. *F. Rouge, Lausanne*, 250, 1896. (Cited on page 50.)

[271] S. Park, D. Lee, and W.W. Chu. Fast retrieval of similar subsequences in long sequence databases. In *In 3rd IEEE Knowledge and Data Engineering Exchange Workshop*, pages 60–67, 1999. (Cited on page 31.)

[272] S. Park, W.W. Chu, J. Yoon, and C. Hsu. Efficient searches for similar subsequences of different lengths in sequence databases. In *Proceedings. 16th International Conference on Data Engineering*, pages 23–32, 2000. ISBN 0769505066. (Cited on page 46.)

[273] P. Patel, E. Keogh, J. Lin, and S. Lonardi. Mining Motifs in Massive Time Series Databases. In *Proceedings of IEEE International Conference on Data Mining (ICDM02)*, pages 370–377, 2002. (Cited on pages 29 and 34.)

[274] G Peeters. A large set of audio features for sound description in the cuidado project. Technical report, IRCAM, Paris, 2004. (Cited on pages 10, 73, and 74.)

[275] G.G. Peeters and E. Deruty. Automatic morphological description of sounds. *Journal of the Acoustical Society of America*, 123(5):3801, 2008. (Cited on page 72.)

[276] A. Perina, M. Cristani, U. Castellani, and V. Murino. A new generative feature set based on entropy distance for discriminative classification. *Image Analysis and Processing–ICIAP 2009*, pages 199–208, 2009. (Cited on pages 93 and 191.)

[277] C.S. Perng, H. Wang, S.R. Zhang, and D.S. Parker. Landmarks : a new model for similarity-based pattern querying in time series databases. In *Proceedings of the 16th International Conference on Data Engineering*, pages 33–42, 2000. (Cited on page 38.)

[278] P. Perny. Multicriteria filtering methods based on concordance and non-discordance principles. *Annals of Operations Research*, 80:137–165, 1998. ISSN 0254-5330. (Cited on page 54.)

[279] M.H. Pesaran, D. Pettenuzzo, and A. Timmermann. Forecasting time series subject to multiple structural breaks. *Review of Economic Studies*, 73(4):1057–1084, 2006. (Cited on page 32.)

[280] K. Phua, T.H. Dat, J. Chen, and L. Shue. Human identification using heart sound. In *Second International Workshop on Multimodal User Authentication*, Toulouse, France, 2006. (Cited on pages 113 and 203.)

[281] K. Phua, J. Chen, T.H. Dat, and L. Shue. Heart sound as a biometric. *Pattern Recognition*, 41(3):906–919, 2008. (Cited on pages 113 and 203.)

[282] W. Piston. *Orchestration*. WW Norton New York, 1955. (Cited on pages 3 and 6.)

[283] R. Plomp. Timbre as a multidimensional attribute of complex tones. *Frequency analysis and periodicity detection in hearing*, pages 397–414, 1970. (Cited on page 15.)

[284] I. Popivanov and R.J. Miller. Similarity search over time-series data using wavelets. In *Proceedings of the International Conference on Data Engineering*, pages 212–224, 2002. (Cited on pages 26 and 36.)

[285] R.J. Povinelli, M.T. Johnson, A.C. Lindgren, and J. Ye. Time series classification using Gaussian mixture models of reconstructed phase spaces. *IEEE Transactions on Knowledge and Data Engineering*, 16(6):779–783, 2004. (Cited on page 30.)

[286] F. Prat, A. Marzal, S. Martin, R. Ramos-Garuo, and M.J. Castro. A template-based recognition system for on-line handwritten characters. *Journal of Information Science and Engineering*, 25(3):779–791, 2009. (Cited on pages 93 and 195.)

[287] J. Pressing. *Synthesizer performance and real-time techniques*. AR Editions, Inc. Madison, WI, USA, 1992. (Cited on page 78.)

[288] D. Pressnitzer, S. McAdams, S. Winsberg, and J. Fineberg. Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Perception and Psychophysics*, 62(1):66–80, 2000. ISSN 0031-5117. (Cited on page 10.)

[289] H. Proença and L.A. Alexandre. The nice. i: noisy iris challenge evaluation-part i. In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pages 1–4. IEEE, 2007. (Cited on page 129.)

[290] D. Psenicka. Sporch: An algorithm for orchestration based on spectral analyses of recorded sounds. In *Proceedings of International Computer Music Conference (ICMC)*, 2003. (Cited on page 149.)

[291] H. Qi, P. Hartono, K. Suzuki, and S. Hashimoto. Sound database retrieved by sound. *Acoustical Science and Technology*, 23(6):293–300, 2002. (Cited on pages 73 and 132.)

[292] M. Radovanović, A. Nanopoulos, and M. Ivanović. Hubs in space: Popular nearest neighbors in high-dimensional data. *The Journal of Machine Learning Research*, 9999:2487–2531, 2010. (Cited on pages 83 and 89.)

[293] M. Radovanovic, A. Nanopoulos, and M. Ivanovic. Time-series classification in many intrinsic dimensions. In *10th SIAM International Conference on Data Mining*. Citeseer, 2010. (Cited on pages 83 and 89.)

[294] D. Rafiei and A. Mendelzon. Efficient Retrieval of Similar Time Sequences Using DFT. In *Proceedings. 5th International Conference of Foundations of Data Organization and Algorithms*, pages 249–257, 1998. (Cited on page 26.)

[295] S. Rahimi, L. Gandy, and N. Mogharreban. A web-based high-performance multicriteria decision support system for medical diagnosis. *International Journal of Intelligent Systems*, 22(10):1083–1099, 2007. ISSN 1098-111X. (Cited on page 54.)

[296] C. Ratanamahatana and D. Wanichsan. Stopping Criterion Selection for Efficient Semi-supervised Time Series Classification. *Studies in Computational Intelligence*, 149:1–14, 2008. (Cited on page 30.)

[297] C. Ratanamahatana, E. Keogh, A.J. Bagnall, and S. Lonardi. A novel bit level time series representation with implication of similarity search and clustering. *Advances in Knowledge Discovery and Data Mining*, pages 771–777, 2005. (Cited on pages 27 and 38.)

[298] C.A. Ratanamahatana and E. Keogh. Everything you know about dynamic time warping is wrong. In *Third Workshop on Mining Temporal and Sequential Data*, pages 1–11, Seattle, WA, USA, 2004. (Cited on page 41.)

[299] C.A. Ratanamahatana and E. Keogh. Making time-series classification more accurate using learned constraints. In *Proceedings of SIAM International Conference on Data Mining*, pages 11–22, 2004. (Cited on page 42.)

[300] G. Ravi, S.K. Gupta, and MB Ray. Multiobjective optimization of cyclone separators using genetic algorithm. *Ind. Eng. Chem. Res*, 39(11):4272–4286, 2000. (Cited on page 54.)

[301] KV Ravi Kanth, D. Agrawal, and A. Singh. Dimensionality reduction for similarity searching in dynamic databases. *ACM SIGMOD Record*, 27(2):166–176, 1998. (Cited on page 36.)

[302] G. Reeves, J. Liu, S. Nath, and F. Zhao. Managing massive time series streams with multi-scale compressed trickles. *Proceedings of the VLDB Endowment*, 2(1): 97–108, 2009. (Cited on pages 35 and 36.)

[303] T. Rehman and C. Romero. The application of the MCDM paradigm to the management of agricultural systems: some basic considerations. *Agricultural Systems*, 41(3):239–255, 1993. ISSN 0308-521X. (Cited on page 54.)

[304] G. Reinert, S. Schbath, and M.S. Waterman. Probabilistic and statistical properties of words: an overview. *Journal of Computational Biology*, 7(1-2):1–46, 2000. (Cited on page 44.)

[305] MJ Reyes-Gomez and DPW Ellis. Selection, parameter estimation, and discriminative training of hidden markov models for general audio modeling. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, volume 1, pages 70–73, 2003. (Cited on pages 133, 138, and 139.)

[306] J.C. Risset. *Hauteur et timbre des sons*. IRCAM, Paris, 1978. (Cited on pages 9 and 11.)

[307] JJ Rodriguez and LI Kuncheva. Time series classification: Decision forests and SVM on interval and DTW features. In *Proc Workshop on Time Series Classification, 13th International Conference on Knowledge Discovery and Data mining*, 2007. (Cited on page 30.)

[308] C. Romero and T. Rehman. Natural resource management and the use of multiple criteria decision-making techniques: A review. *European Review of Agricultural Economics*, 14(1):61, 1987. ISSN 0165-1587. (Cited on page 54.)

[309] F. Rose and J. Hetrick. Spectral analysis as a ressource for contemporary orchestration technique. In *Proceedings of Conference on Interdisciplinary Musicology*, volume 2005, 2005. (Cited on page 149.)

[310] Y. Sakurai, M. Yoshikawa, S. Uemura, and H. Kojima. The A-tree: An index structure for high-dimensional spaces using relative approximation. In *Proceedings of the 26th International Conference on Very Large Data Bases*, pages 516–526, 2000. (Cited on page 46.)

[311] Y. Sakurai, M. Yoshikawa, and C. Faloutsos. FTW: fast similarity search under the time warping distance. In *Proceedings of the 24th ACM Symposium on Principles of database systems*, pages 326–337, 2005. (Cited on page 27.)

[312] S. Salvador and P. Chan. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580, 2007. (Cited on page 42.)

[313] S. Salvador, P. Chan, and J. Brodie. Learning states and rules for time series anomaly detection. In *Proc. 17th International FLAIRS Conference*, pages 300–305, 2004. (Cited on page 33.)

[314] S.L. Salzberg. On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Mining and knowledge discovery*, 1(3):317–328, 1997. (Cited on pages 92, 94, and 96.)

[315] S.L. Salzberg. On comparing classifiers: A critique of current research and methods. *Data mining and knowledge discovery*, 1(1), 1999. (Cited on pages 92 and 94.)

[316] S. Samson, R.J. Zatorre, and J.O. Ramsay. Multidimensional scaling of synthetic musical timbre: Perception of spectral and temporal characteristics. *Canadian Journal of Experimental Psychology*, 51(4):307–315, 1997. ISSN 1196-1961. (Cited on page 15.)

[317] P. Schaeffer. *Traité des objets musicaux*. Seuil, 1966. (Cited on pages 5, 7, and 139.)

[318] S. Schliebs, H. Hamed, and N. Kasabov. Reservoir-based evolving spiking neural network for spatio-temporal pattern recognition. In *Neural Information Processing*, pages 160–168. Springer, 2011. (Cited on pages 93 and 194.)

[319] JF Schouten. The perception of timbre. In *Reports of the 6th International Congress on Acoustics*, volume 76, 1968. (Cited on page 11.)

[320] T.B. Sebastian, P.N. Klein, and B.B. Kimia. On aligning curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1):116–125, 2003. (Cited on page 49.)

[321] P. Sebastiani, M. Ramoni, P. Cohen, J. Warwick, and J. Davis. Discovering dynamics using Bayesian clustering. *Lecture Notes in Computer Science*, 1642: 199–209, 1999. (Cited on pages 39 and 44.)

[322] J. Seibert. Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earth System Sciences*, 4(2):215–224, 2000. (Cited on page 54.)

[323] J. Seinfeld and W. McBride. Optimization with multiple performance criteria. *Ind. Eng. Chem. Process Des. Develop*, 9(1):53–57, 1970. (Cited on page 54.)

[324] A. Sfetsos and C. Siriopoulos. Time series forecasting with a hybrid clustering scheme and pattern recognition. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 34(3):399–405, 2004. (Cited on page 32.)

[325] X. Shao, C. Xu, and M.S. Kankanhalli. Applying neural network on the content-based audio classification. In *Proceedings of the 4th IEEE Joint International Conference on Information, Communications and Signal Processing*, volume 3, pages 1821–1825, 2003. (Cited on pages 133 and 138.)

[326] D.E. Shasha and Y. Zhu. *High performance discovery in time series: techniques and case studies*. Springer-Verlag New York Inc, 2004. (Cited on pages 31 and 36.)

[327] H. Shatkay and SB Zdonik. Approximate queries and representations for large data sequences. In *Data Engineering, 1996. Proceedings of the Twelfth International Conference on*, pages 536–545, 1996. (Cited on pages 26, 30, 38, and 43.)

[328] J. Shieh and E. Keogh. isax : indexing and mining terabyte sized time series. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 623–631. ACM, 2008. (Cited on pages 38, 39, 47, and 62.)

[329] V. G. Sigillito, S. P. Wing, L. V. Hutton, and K. B. Baker. Classification of radar returns from the ionosphere using neural networks. *Johns Hopkins APL Technical Digest*, pages 262–266, 1989. (Cited on page 194.)

[330] Y.N. Singh and SK Singh. Evaluation of electrocardiogram for biometric authentication. *Journal of Information Security*, 3:39–48, 2012. (Cited on page 129.)

[331] M. Slaney. Mixtures of probability experts for audio retrieval and indexing. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, volume 1, pages 345–348. IEEE, 2002. (Cited on page 71.)

[332] M. Slaney. Semantic-audio retrieval. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 4, pages IV–4108, 2002. (Cited on page 71.)

[333] P. Smyth. Clustering sequences with hidden Markov models. *Advances in Neural Information Processing Systems*, pages 648–654, 1997. (Cited on page 28.)

[334] H. Song and G. Li. Tourism demand modelling and forecasting–A review of recent research. *Tourism Management*, 29(2):203–220, 2008. (Cited on pages 23 and 32.)

[335] A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, and A. Lendasse. Methodology for long-term prediction of time series. *Neurocomputing*, 70(16-18):2861–2869, 2007. (Cited on page 32.)

[336] D. Srisai and C.A. Ratanamahatana. Efficient Time Series Classification under Template Matching Using Time Warping Alignment. In *Proceedings of the Fourth International Conference on Computer Sciences and Convergence Information Technology*, pages 685–690. IEEE, 2009. (Cited on page 30.)

[337] T. Stiefmeier, D. Roggen, and G. Troster. Gestures are strings: Efficient online gesture spotting and classification using string matching. In *Proceedings of the ICST 2nd international conference on Body area networks*, pages 1–8, Florence, Italy, 2007. (Cited on pages 27, 35, and 36.)

[338] RG Stockwell, L. Mansinha, and RP Lowe. Localization of the complex spectrum: the s transform. *Signal Processing, IEEE Transactions on*, 44(4):998–1001, 1996. (Cited on pages 117 and 204.)

[339] ZR Struzik, A. Siebes, and A. CWI. Measuring time series similarity through large singular features revealed with wavelet transformation. In *Proceedings of the Tenth International Workshop on Database and Expert Systems Applications*, pages 162–166, 1999. (Cited on page 36.)

[340] A. Subasi. EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, 32(4):1084–1093, 2007. (Cited on page 30.)

[341] SR Subramanya and A. Youssef. Wavelet-based indexing of audio data in audio/-multimedia databases. In *Multi-Media Database Management Systems, 1998. Proceedings. International Workshop on*, pages 46–53. IEEE, 1998. (Cited on page 132.)

[342] SR Subramanya, R. Simha, B. Narahari, and A. Youssef. Transform-based indexing of audio data for multimedia databases. In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems' 97*, pages 211–218. IEEE, 1997. (Cited on pages 71 and 132.)

[343] F. Sufi, I. Khalil, and J. Hu. Ecg-based authentication. *Handbook of Information and Communication Security*, pages 309–331, 2010. (Cited on page 113.)

[344] S. Sundaram and S. Narayanan. Audio retrieval by latent perceptual indexing. In *Acoustics, Speech and Signal Processing. ICASSP 2008. IEEE International Conference on*, pages 49–52. IEEE, 2008. (Cited on page 72.)

[345] J. Sundberg. Musical significance of musicians syllable choice in improvised nonsense text singing: A preliminary study. *Phonetica*, 51(1-3):132–145, 1994. (Cited on page 78.)

[346] J. Sundberg. Level and Center Frequency of the Singer's Formant* 1. *Journal of voice*, 15(2):176–186, 2001. (Cited on page 78.)

[347] N. Takama and T. Umeda. Multi-level, multi-objective optimization in process engineering. *Chemical Engineering Science*, 36(1):129–136, 1981. ISSN 0009-2509. (Cited on page 54.)

[348] P. Tan and D. Dowe. Mml inference of oblique decision trees. *AI 2004: Advances in Artificial Intelligence*, pages 321–338, 2005. (Cited on pages 93 and 199.)

[349] H. Tang and S.S. Liao. Discovering original motifs with different lengths from time series. *Knowledge-Based Systems*, 21(7):666–671, 2008. (Cited on page 34.)

[350] D. Tardieu. *Modèles d'instruments pour l'aide à l'orchestration*. PhD thesis, 2008. (Cited on page 150.)

[351] D. Tardieu and X. Rodet. An instrument timbre model for computer aided orchestration. In *Applications of Signal Processing to Audio and Acoustics, 2007 IEEE Workshop on*, pages 347–350. IEEE, 2007. (Cited on pages 150, 152, and 153.)

[352] D. Tardieu, G. Carpentier, and X. Rodet. Computer-aided orchestration based on probabilistic instruments models and genetic exploration. In *The Proceedings of International Computer Music Conference*, pages 188–91, 2007. (Cited on page 150.)

[353] T. Theodoridis and H. Hu. Action classification of 3d human models using dynamic anns for mobile robot surveillance. In *Robotics and Biomimetics, 2007. ROBIO 2007. IEEE International Conference on*, pages 371–376. IEEE, 2007. (Cited on pages 93, 196, and 200.)

[354] T.A. Traill and N.J. Fortuin. Presystolic mitral closure sound in aortic regurgitation with left ventricular hypertrophy and first degree heart block. *British Heart Journal*, 48(1):78, 1982. (Cited on page 113.)

[355] D.H. Tran, Y.R. Leng, and H. Li. Feature integration for heart sound biometrics. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 1714–1717. IEEE, 2010. (Cited on page 204.)

[356] R.S. Tsay. *Analysis of financial time series*. Wiley-Interscience, 2005. ISBN 0471690740. (Cited on page 31.)

[357] A. Tsoukas, M. Tirrell, and G. Stephanopoulos. Multiobjective dynamic optimization of semibatch copolymerization reactors. *Chemical Engineering Science*, 37(12): 1785–1795, 1982. ISSN 0009-2509. (Cited on page 54.)

[358] A. Uitdenbogerd and J. Zobel. Melodic matching techniques for large music databases. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, page 66. ACM, 1999. (Cited on page 132.)

[359] M. Van Gerven and O. Jensen. Attention modulations of posterior alpha as a control signal for two-dimensional brain-computer interfaces. *Journal of neuroscience methods*, 179(1):78–84, 2009. (Cited on pages 93 and 190.)

[360] D.A. van Leeuwen, A.F. Martin, M.A. Przybocki, and J.S. Bouten. Nist and nfi-tno evaluations of automatic speaker recognition. *Computer Speech & Language*, 20(2): 128–158, 2006. (Cited on page 129.)

[361] A. Van Oosterom, R. Hoekema, and GJH Uijen. Geometrical factors affecting the interindividual variability of the ecg and the vcg. *Journal of electrocardiology*, 33: 219–227, 2000. (Cited on page 131.)

[362] K.T. Vasko and H.T.T. Toivonen. Estimating the number of segments in time series data using permutation tests. In *Proceedings of the IEEE International Conference on Data Mining*, pages 466–473, 2002. (Cited on page 31.)

[363] T. Virtanen and M. Helen. Probabilistic model based similarity measures for audio query-by-example. In *Applications of Signal Processing to Audio and Acoustics, 2007 IEEE Workshop on*, pages 82–85, 2007. (Cited on pages 15, 57, 72, and 133.)

[364] M. Vlachos, D. Gunopoulos, and G. Kollios. Discovering similar multidimensional trajectories. In *Proceedings of the 18th International Conference on Data Engineering*, pages 673–684. IEEE Computer Society, 2002. (Cited on pages 27, 35, and 42.)

[365] M. Vlachos, J. Lin, E. Keogh, and D. Gunopulos. A wavelet-based anytime algorithm for k-means clustering of time series. In *Proc. Workshop on Clustering High Dimensionality Data and Its Applications*, pages 23–30, 2003. (Cited on page 28.)

[366] M. Vlachos, D. Gunopulos, and G. Das. Indexing time-series under conditions of noise. *Data mining in time series databases*, pages 67–100, 2004. (Cited on pages 27 and 36.)

[367] M. Vlachos, P. Yu, and V. Castelli. On periodicity detection and structural periodic similarity. In *SIAM International Conference on Data Mining*, pages 449–460, Newport Beach, CA, 2005. (Cited on page 43.)

[368] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos, and E. Keogh. Indexing multi-dimensional time-series. *The VLDB Journal*, 15(1):1–20, 2006. (Cited on pages 42 and 47.)

[369] G. Von Bismarck. Sharpness as an attribute of the timbre of steady sounds. *Acustica*, 30(3):159–172, 1974. (Cited on page 10.)

[370] N. Wagner, Z. Michalewicz, M. Khouja, and R.R. McGregor. Time series forecasting for dynamic environments: the DyFor genetic program model. *IEEE transactions on evolutionary computation*, 11(4):433–452, 2007. (Cited on page 32.)

[371] C. Wan and M. Liu. Content-based audio retrieval with relevance feedback. *Pattern recognition letters*, 27(2):85–92, 2006. (Cited on page 132.)

[372] C. Wan, M. Liu, and L. Wang. Content-based sound retrieval for web application. *Web Intelligence: Research and Development*, pages 389–393, 2001. (Cited on pages 71 and 132.)

[373] Y. Wang, F. Agrafioti, D. Hatzinakos, and K.N. Plataniotis. Analysis of human electrocardiogram for biometric recognition. *EURASIP journal on Advances in Signal Processing*, 2008:19, 2008. (Cited on page 113.)

[374] AS Weigend and NA Gershenfeld. *Time Series Prediction: forecasting the future and understanding the past*. Addison Wesley, 1994. (Cited on page 23.)

[375] G.M. Weiss. Mining with rarity: a unifying framework. *ACM SIGKDD Explorations Newsletter*, 6(1):7–19, 2004. (Cited on pages 23 and 33.)

[376] D.L. Wessel. Timbre space as a musical control structure. *Computer music journal*, 3(2):45–52, 1979. ISSN 0148-9267. (Cited on page 15.)

[377] G. Wichern, J. Xue, H. Thornburg, and A. Spanias. Distortion-aware query-by-example for environmental sounds. In *Applications of Signal Processing to Audio and Acoustics, 2007 IEEE Workshop on*, pages 335–338, 2007. (Cited on page 72.)

[378] B. Williams, M. Toussaint, and A. Storkey. Extracting motion primitives from natural handwriting data. *Artificial Neural Networks–ICANN 2006*, pages 634–643, 2006. (Cited on page 191.)

[379] B. Williams, M. Toussaint, and A. Storkey. Modelling motion primitives and their timing in biologically executed movements. *Advances in Neural Information Processing Systems*, 20:1609–1616, 2008. (Cited on page 191.)

[380] E. Wold, T. Blum, D. Keislar, and J. Wheaten. Content-based classification, search, and retrieval of audio. *Multimedia, IEEE*, 3(3):27–36, 1996. (Cited on pages 71, 132, 133, 138, and 139.)

[381] D.H. Wolpert and W.G. Macready. No free lunch theorems for optimization. *Evolutionary Computation, IEEE Transactions on*, 1(1):67–82, 1997. (Cited on page 96.)

[382] X. Xi, E. Keogh, C. Shelton, L. Wei, and C.A. Ratanamahatana. Fast time series classification using numerosity reduction. In *Proceedings of the 23rd international conference on Machine learning ICML 06*, volume 150, pages 1033–1040. ACM Press, 2006. (Cited on pages 30, 89, and 95.)

[383] X. Xi, E. Keogh, L. Wei, and A. Mafra-neto. Finding Motifs in Database of Shapes. In *Proc. of SIAM International Conference on Data Mining*, pages 249–260, Minneapolis, Minnesota, USA, 2007. (Cited on page 49.)

[384] H. Xia, G.A. Garcia, J.C. McBride, A. Sullivan, T. De Bock, J. Bains, D.C. Wortham, and X. Zhao. Computer algorithms for evaluating the quality of ecgs in real time. 2012. (Cited on pages 93 and 190.)

[385] J. Xie and WY Yan. Pattern-based characterization of time series. *International Journal of Information and Systems Science*, 3(3):479–491, 2007. (Cited on page 38.)

[386] Y. Xiong and D.Y. Yeung. Time series clustering with ARMA mixtures. *Pattern Recognition*, 37(8):1675–1689, 2004. (Cited on page 43.)

[387] J. Xue, G. Wichern, H. Thornburg, and A. Spanias. Fast query by example of environmental sounds via robust and efficient cluster-based indexing. In *Acoustics, Speech and Signal Processing. ICASSP 2008. IEEE International Conference on*, pages 5–8, 2008. (Cited on page 72.)

[388] RN Yadav, PK Kalra, and J. John. Time series prediction with single multiplicative neuron model. *Applied soft computing*, 7(4):1157–1163, 2007. (Cited on page 32.)

[389] N. Yager and T. Dunstone. The biometric menagerie. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(2):220–230, 2010. (Cited on page 126.)

[390] D. Yankov, E. Keogh, J. Medina, B. Chiu, and V. Zordan. Detecting time series motifs under uniform scaling. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 844–853. ACM, 2007. (Cited on page 34.)

[391] D. Yankov, E. Keogh, and U. Rebbapragada. Disk aware discord discovery: Finding unusual time series in terabyte sized datasets. *Knowledge and Information Systems*, 17(2):241–262, 2008. (Cited on page 33.)

[392] P.O. Yapo, H.V. Gupta, and S. Sorooshian. Multi-objective global optimization for hydrologic models. *Journal of hydrology*, 204(1-4):83–97, 1998. ISSN 0022-1694. (Cited on page 54.)

[393] D. Ye, X. Wang, E. Keogh, and A. Mafra-Neto. Autocannibalistic and Anyspace Indexing Algorithms with Applications to Sensor Data Mining. In *The SIAM International Conference on Data Mining (SDM 2009)*, pages 85–96, Sparks, Nevada, 2009. (Cited on page 48.)

[394] L. Ye and E. Keogh. Time series shapelets: a new primitive for data mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 947–956. ACM, 2009. (Cited on pages 38 and 49.)

[395] B.K. Yi and C. Faloutsos. Fast time sequence indexing for arbitrary Lp norms. In *Proceedings of the 26th International Conference on Very Large Data Bases*, pages 385–394, 2000. (Cited on pages 26, 36, 39, 40, and 47.)

[396] B.K. Yi, HV Jagadish, and C. Faloutsos. Efficient retrieval of similar time sequences under time warping. In *Data Engineering, 1998. Proceedings., 14th International Conference on*, pages 201–208, 1998. (Cited on page 27.)

[397] H. Yoon, K. Yang, and C. Shahabi. Feature subset selection and feature ranking for multivariate time series. *IEEE transactions on knowledge and data engineering*, pages 1186–1198, 2005. (Cited on page 28.)

[398] A. Yoshitaka and T. Ichikawa. A survey on content-based retrieval for multimedia databases. *IEEE Transactions on Knowledge and Data Engineering*, 11(1):81–93, 1999. (Cited on page 71.)

[399] A. Ypma and R.P.W. Duin. Novelty detection using self-organizing maps. *Progress in Connectionist-Based Information Systems*, 2:1322–1325, 1997. (Cited on page 33.)

[400] CC Yuen, SK Gupta, and AK Ray. Multi-objective optimization of membrane separation modules using genetic algorithm. *Journal of Membrane Science*, 176(2): 177–196, 2000. ISSN 0376-7388. (Cited on page 54.)

[401] MF Zafar, D. Mohamad, and MM Anwar. Recognition of online isolated handwritten characters by backpropagation neural nets using sub-character primitive features. In *Multitopic Conference, 2006. INMIC'06. IEEE*, pages 157–162. IEEE, 2006. (Cited on page 191.)

[402] Y.Y. Zhan, X.Y. Chen, and R.C. Xu. Outlier detection algorithm based on pattern representation of time series. *Application Research of Computers*, 24(11):96–99, 2007. (Cited on page 38.)

[403] T. Zhang and C.C.J. Kuo. Classification and retrieval of sound effects in audiovisual data management. In *Signals, Systems, and Computers, 1999. Conference Record of the Thirty-Third Asilomar Conference on*, volume 1, pages 730–734. IEEE, 1999. (Cited on pages 71 and 132.)

[404] T. Zhang and C.C.J. Kuo. Hierarchical classification of audio data for archiving and retrieving. In *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing*, pages 3001–3004, 1999. (Cited on page 72.)

[405] X. Zhang, J. Wu, X. Yang, H. Ou, and T. Lv. A novel pattern extraction method for time series classification. *Optimization and Engineering*, 10(2):253–271, 2009. (Cited on pages 30 and 34.)

[406] X.L. Zhang, H. Begleiter, B. Porjesz, W. Wang, and A. Litke. Event related potentials during object recognition tasks. *Brain Research Bulletin*, 38(6):531–538, 1995. (Cited on page 191.)

[407] Z. Zhang, Y. Shi, and G. Gao. A rough set-based multiple criteria linear programming approach for the medical diagnosis and prognosis. *Expert Systems with Applications*, 36(5):8932–8937, 2009. ISSN 0957-4174. (Cited on page 54.)

[408] S. Zhong and J. Ghosh. HMMs and coupled HMMs for multi-channel EEG classification. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, pages 1154–1159, 2002. (Cited on pages 30, 93, and 192.)

[409] S. Zhong, T.M. Khoshgoftaar, and N. Seliya. Clustering-based network intrusion detection. *International Journal of Reliability Quality and Safety Engineering*, 14(2): 169–187, 2007. (Cited on pages 23, 29, and 33.)

[410] Y. Zhu and D. Shasha. Query by humming: a time series database approach. In *Proc. of SIGMOD*, page 675, 2003. (Cited on page 71.)

[411] E. Zitzler, L. Thiele, M. Laumanns, C.M. Fonseca, and V.G. da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *Evolutionary Computation, IEEE Transactions on*, 7(2):117–132, 2003. (Cited on page 84.)