Computer Aided Design of Audio Signal Classifier Systems

Sebastian Huebner

Pestalozzistrasse 5, 14482 Potsdam, Germany
sebastian@sejona.de

In the article an approach for computer aided design (CAD) of audio signal classifiers is described. Outlined is the basic architecture of a general purpose CAD framework for the bioacoustics domain. Emphasis is placed on both the role of expert knowledge and of machine learning in the classifier design process.

1 Introduction

For the end user most systems for bioacoustic pattern recognition are black boxes. Researchers usually cannot understand how these systems work, why classification decisions (annotations) are generated and why false decisions occur. This is not surprising: Automated classification of acoustic patterns is a difficult computational task. Algorithms are just as complicated as in other fields of pattern recognition, e.g. in picture processing. Knowledge in mathematics and in computer science is necessary to understand how they work.

An alternative to traditional black box sound recognition systems can be an environment for computer aided design (CAD) of audio signal classifiers. The purpose of such an environment is to facilitate modeling, implementation, test and application of audio classifier systems and to hide from the end user the complexity of all involved digital signal processing and pattern recognition algorithms. In this article theory and fundamental architecture of such an environment are described. Practical examples can be found on the author's homepage.

2 CAD for audio signal classifiers

In order to understand the reasons that speak for a CAD approach in the area of bioacoustic pattern recognition it is best to first have a look at the advantages and disadvantages of the traditional or "black box" approach. Afterwards, it will become evident that a CAD approach is superior in this highly knowledge-intensive domain.

2.1 The black box approach

The task of a black box classifier system (see figure 1) is to process bioacoustic signals and to generate annotations (classification decisions) that indicate the presence of certain acoustic event types at certain points of time in the audio data.

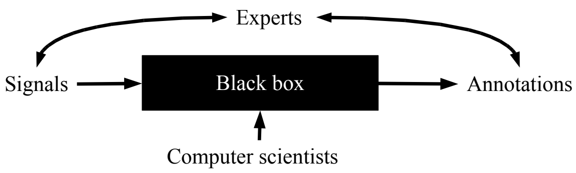

Figure 1: The black box approach

In many cases it is the explicit wish of both the computer scientists (the constructors of the black box) and the experts (biologists and field researchers) that classification is done unsupervised by the black box.

Usually, after the box has been implemented, the job of the computer scientists is done. Expert's control on the system is restricted to the input (audio signals) and possibly to a few general parameters that configure algorithms inside the black box. The output (annotations) can neither be influenced nor be understood properly by the experts. This is especially awkward if within an unsupervised classification process many false classification decisions occur for unknown reasons. Experts do not have the slightest chance to correct this type of ill behavior even if the underlying problem actually is trivial.

It may be said that this approach fails the more often, the more complex the classification task is. It may also be said that the more complex the classification task is, the more expertise in bioacoustics and biology is needed to build the black box itself. Indeed, to solve complex bioacoustic classification problems within a black box approach a close cooperation of computer scientists and biologists is inevitable.

In many cases experts start understanding the nature of specific bioacoustic signals only after having intensively worked with them for a while. Expertise grows slowly but is needed to build and modify the black box. Unfortunately, a modification of the box requires each time a close cooperation with the computer scientists. In order to take new important scientific insights into account, they have to be engaged for one more time.

It should be mentioned that the black box approach has also some advantages. A black box can be a very compact efficiently implemented solution. Bioacousticians don't have to deal with computational details and may entirely rely on the qualifications of the programmers. However, the above mentioned circumstances lead to a whole series of severe disadvantages characteristic for the black box approach:

- Experts have low influence on how their expertise is implemented.
- Expertise incorporated in classifier systems is not sharable among experts.
- Computer scientist's work is difficult to re-use.
- Experts can't understand why the system behaves the way it does.
- Experts have little chance to correct an ill-behaving system.
- The incorporation of new scientific insights into the nature of bioacoustic signals requires a re-implementation of parts of the system.
- General inflexibility and expensiveness.

In many other expertise-intensive problem domains interactive CAD solutions proved to be much advantageous compared to black boxes. How CAD for audio signal classifiers can be understood, is described in the next section.

2.2 The CAD approach

Computer aided knowledge design (knowledge engineering) is the art and science to transfer domain specific expertise to a computer system in such a way that the expertise can both be computed by the system and results still be understood by human experts. The claims of computability and at the same time comprehensibility require a new conception of acoustic pattern recognition systems. A **c**omputer **a**ided **d**esign system for **a**udio **c**lassifiers (CADAC) is different from the above described black box solutions (see figure 2):

1. The CADAC is able to annotate audio material only if equipped with a library of previously designed classifiers.

2. The CADAC is made to give experts (non-technicians) the opportunity to create their own classifier libraries by making use of their individual expertise. The creation of such libraries includes prototyping, modeling, implementation, test and application of classifiers.

3. The CADAC is equipped with a special CAD interface. This interface provides all means to model, test and apply classifiers for audio data and at the same time hides the complexity of the machinery that carries out difficult computations.

4. The CADAC guarantees full transparency on all scientifically important levels to the experts. Subject of computation are not nebulous concepts, obscure networks or sophisticated but incomprehensible systems but well defined signatures of acoustic event types and comprehensible similarity measures.

5. The CADAC provides immediate visual feedback within all steps of the classifier design process. Visual feedback concerns the structure of signatures of acoustic events, similarity measuring functions and classifier decisions generated by the system.

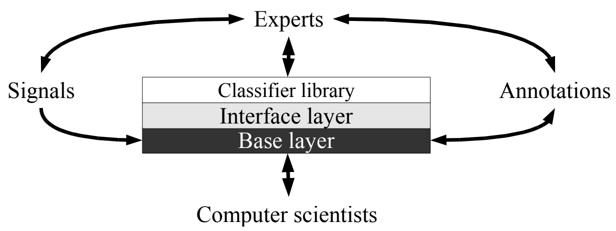

Figure 2: The CAD approach

It is easy to see that such an environment is far more complex than just a black box classification engine made to respond to certain patterns in audio signals. Especially the CAD interface is an extensive and complicated piece of software though at first glance it appears trivial as it's main purpose is to solve pragmatic problems.

However, given a state of the art CAD interface the architecture shown in figure 2 opens new possibilities. High performance pattern recognition algorithms in the base layer may be used by experts interactively and independently to create and modify classifiers in accordance with their specific needs. New scientific insights may be taken into account immediately by the experts.

The CAD approach has also some disadvantages. The underlying software is more complex and experts have to do design work on their own. Also the needs of full transparency and comprehensibility within all steps of the classifier design process may cause problems with certain types of existing classification algorithms. However, the above described architecture leads to a whole series of advantages:

- Experts have full control on the interactive classifier design process.
- Experts can understand what is computed and why annotations are generated.
- Expertise incorporated in classifiers is sharable worldwide.
- Computer scientist's work (the CADAC itself) is suitable for a virtually infinite number of different tasks.
- Experts can easily correct an ill-behaving system (even if the underlying problem is not trivial).
- The incorporation of new insights into the nature of acoustic signals does not necessarily require reimplementations of parts of the system.
- Knowledge discovery and data mining instruments easily fit into the architecture.
- General flexibility and cheapness in the long term.

It can be concluded that the CAD approach does not only not suffer from the disadvantages of the black box approach but has several additional advantages.

3 Architecture

A CADAC for bioacoustic classifier design comprises a large variety of tools more or less independent from one another. The most essential of them may be grouped into three layers (see figure 2). In figure 3 these layers are explained in more detail. Small helper-tools are not included in the description though they also play an important role in practical design tasks.

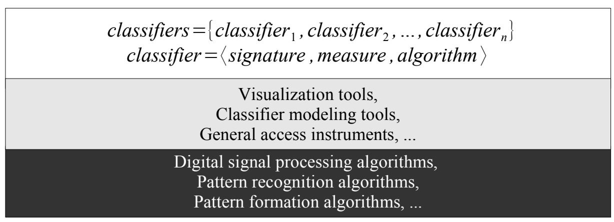

Figure 3: Layers of a CADAC

3.1 The classifier library

Transparency and comprehensibility are the two most important features of the classifiers in the library. To achieve this, a classifier can be defined as a triple consisting of a *signature*, a *similarity measure* and an *annotation generating algorithm*:

1. The *signature* is a well defined descriptor of the class of acoustic events that the classifier is supposed to detect in audio signals. Expertise is contained mainly in the signature - not in the other two constituents of the classifier. Signatures may be either subsymbolic or symbolic.

2. The *similarity measure* is a well defined function that is suitable to compare a signature with a piece of audio data of exactly the same duration. By applying signature and similarity measure to a longer audio signal it is possible to compute the degree of similarity between signature and signal for virtually each point of time.

3. The *annotation generating algorithm* systematically evaluates an audio signal by using only signature and similarity measure. The algorithm creates annotations when the degree of computed similarity is high and certain additional constraints are satisfied. Annotations optionally may include additional automatically measured information.

The above described classifier architecture lies at the heart of a comprehensible CADAC. Shapeliness of both signature and similarity measure as well as purity of the annotation generating algorithm are decisive for scientific quality of both design and classification processes. Furthermore, properties and behavior of all three classifier constituents may easily be visualized and thus understood immediately by experts.

3.2 The CAD interface

Classifiers for the library have to be designed in accordance with the requirements of the study or application they are needed for. The CAD interface gives experts the opportunity to model classifiers independently from computer experts. This layer may be grouped into visualization, classifier modeling and general access tools.

1. *Visualization tools* serve three purposes: (1) Visualization of audio data (e.g. in form of spectrograms), (2) visualization of classifiers (including signatures, numerical and structural properties) and (3) visualization of classifier decisions (including the behavior of similarity measuring functions). Visualization tools need to be both intuitive and precise.

2. *Classifier modeling tools* serve all tasks necessary to initialize and modify classifiers and their constituents. Manual editing tools known from picture editors belong to this group as well as automatic signature extraction, merging and clustering tools.

3. *General access tools* are necessary to control all data flows within the CADAC. They serve the access to audio file collections, sets of annotations as well as the configuration and conduction of classification tasks. Access instruments can have the form of wizards that guide through different procedures.

A CAD interface may include very many different tools. For example, signature extraction can be based on dozens of different time-frequency or time-energy based visualizations. For this reason the interface has to be modular and open.

3.3 The base layer

The foundation of a CADAC is the base layer. It is invisible to experts. For them it is almost a black box but it's algorithms may be freely accessed through the CAD interface. Tools in this layer may be grouped into digital signal processing, pattern recognition and pattern formation algorithms.

1. *Digital signal processing algorithms* are necessary for all kinds of computations typical for audio data processing. Included in this group are time-energy, time-frequency an wavelet transforms as well as filtering, signal generating and mapping algorithms.

2. *Pattern recognition algorithms* are used to compute annotations by searching patterns in audio signals. Annotation generating algorithms actually are physically represented in this layer, not inside the classifiers.

3. *Pattern formation algorithms* enhance the CADAC with data mining functionality. They are needed for inductive programming of signatures on subsymbolic and symbolic levels. Note that inductive inference of signatures is one of the most promising applications in the given context.

Today, we can choose from a variety of high performance algorithms for the first two subgroups of the base layer. Most advanced are DSP standard algorithms like DFT or Wavelet-transform computing routines. Selected pattern recognition algorithms may be adopted from natural language and picture processing. Suitable pattern formation algorithms, however, are rare - especially in the field of subsymbolic signature formation. Symbolic data mining methods can be used to work with sets of annotations.

3.4 The classifier design process

Knowledge engineering simple classifiers can be a straightforward task. In such cases the classifier design process fits into the waterfall-scheme known from traditional software engineering. This scheme however is applicable only if the structure of acoustic events is deterministic and *a priori* known to the expert who designs the classifiers.

In some cases CAD of complex classifiers or classifier systems can be seen as a *knowledge discovery* task. In

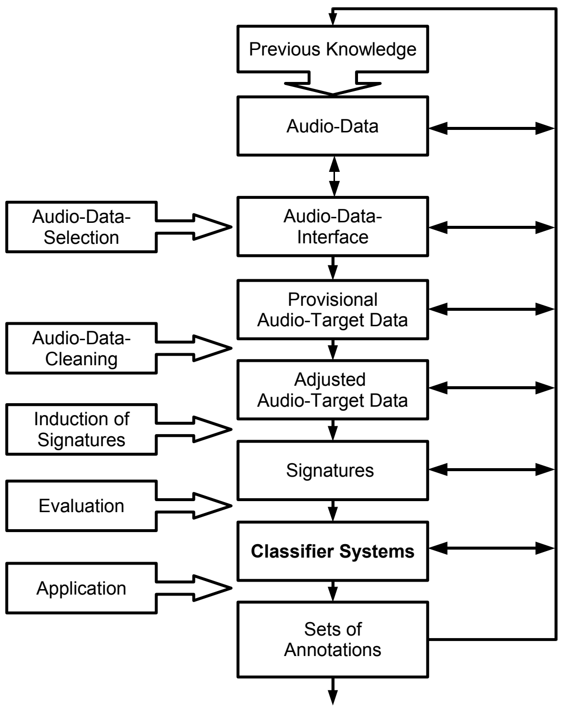

Figure 4: KDD for audio signal classifier design

such cases the classifier design process should be conducted within a KDD-process (see figure 4) [3]. This scheme is applicable if either the structure of acoustic events is not deterministic or not known *a priori* to the expert.

The conduction of a KDD-process on audio file collections can lead not only to proper classifier libraries but also to the discovery of new and interesting acoustic patterns. Such patterns can become relevant to scientific discoveries.

4 Conclusion

Computer aided design of audio signal classifiers is a dynamic and interactive process. Unfortunately, both dynamics and interacitivity cannot be transferred to two dimensional paper. Please visit author's homepage: `http://www.sejona.de` for demos and examples.

References

[1] S. V. Huebner, "Wissensbasierte Modellierung von Audio-Signal-Klassifikatoren. Zur Bioakustik von *Tursiops truncatus.*", *Dissertation, University of Potsdam*, (2007)

[2] S. V. Huebner, "Wissensbasierte Modellierung von Klassifikatoren fue Zeit-Frequenz-Muster in PCM-Daten.", *Logos, Berlin*, (2007)

[3] G. Shapiro, W. Frawley, "Knowledge Discovery in Databases", *AAAI/MIT Press, Cambridge, MA*, (1991)