Managing Animal Sounds - Some Challenges and Research Directions

Daniel Cintra Cugler[1], **Claudia Bauzer Medeiros**[1], **Luís Felipe Toledo**[2]

[1]Institute of Computing
University of Campinas – UNICAMP
13.083-970 – Campinas – SP – Brazil

[2]Museum of Zoology "Prof. Adão José Cardoso", Institute of Biology
University of Campinas – UNICAMP
13.083-862 – Campinas – SP – Brazil

{danielcugler, cmbm}@ic.unicamp.br, toledolf2@yahoo.com.br

Abstract. *For decades, biologists around the world have recorded animal sounds. As the number of records grows, so does the difficulty to manage them, presenting challenges to save, retrieve, share and manage the sounds. This paper presents our preliminary results concerning management of large volumes of animal sound data. The paper also provides an overview from our prototype, an online environment focused on management of this data. This paper also discusses our case study, concerning more than 1 terabyte of animal recordings from Fonoteca Neotropical "Jacques Vielliard", at UNICAMP, Brazil.*

1. Introduction

Earlier animal recordings were commonly stored in magnetic tapes. These recordings are of priceless value, and therefore, special attention must be spent in order to maintain these media clean, free of humidity and fungus infection. More recently, recordings use devices that save data in digital format, such as ATRAC, WAV and MP3.

One initial hurdle is data conversion. In UNICAMP, for instance, researchers have amassed the largest collection of animal recordings in the Neotropics. All of the records are offline, and only recently has there been a concentrated effort in creating digital metadata files for the recordings and converting them to digital media - a painstaking work from expert curators. Converting magnetic tapes to digital format is just part of the problem.

A major challenge is how to conveniently store the data so that other scientists can use them [Gorder 2006]. Besides archival issues, there are many directions that need to be considered in managing animal sound collections. Some such challenges concern the target domain of biology studies - e.g. biologists need to analyze and compare several physical characteristics of the recordings, such as frequency (lower, higher, dominant), call/note/pulse duration, number of pulsers/note, number of notes/call and number of harmonics. Other challenges involve engineering issues - e.g., spatial sound systems, analysis, classification and separation of sounds, automatic speech recognition or even audio compression [Cobos and Lopez 2010].

Still other challenges are directly related to sound management – e.g., data structures, indexing, retrieval, querying and mining. Additional topics involve content-based retrieval (e.g., the input to a query is a sound recording), or web-based data management (to allow interoperability across distinct collections). This whole scenario is typical of eScience research, in which many kinds of scientists need to get together to advance in their fields.

When one considers possible scenarios of research involving animal recordings, it is possible to define at least two groups of interest:

- Biologists – to study countless aspects of animal communication
- Citizens – to identify animal sounds, or just learn to recognize them

In this context, this paper discusses preliminary research results involving researchers from the Institutes of Computing and Biology at UNICAMP, concerning challenges on managing animal recordings, specially birds. Section 2 discusses related work. Section 3 describes features and main challenges faced on starting to digitalize bird sounds from Fonoteca Neotropical "Jacques Vielliard". Section 4 shows a prototype developed to manage sound data. Section 5 presents final considerations and ongoing work.

2. Related Work and Some Challenges

There are several research directions concerning sound management, as mentioned in the introduction. This section concentrates on issues related to sound databases, and some systems developed to manage animal sounds.

First of all, sound is a wave, and as such there have been studies to perform sound mining – e.g., [Weihs et al. 2007]. As pointed out by [Gerhardt and Huber 2002], sound representation is at least bidimensional, requiring considering pitch and duration, and thus traditional series mining techniques must be extended. Similarly, [Uitdenbogerd and Zobel 1999] propose an algorithm for query collections of MIDI files, searching music by similarity. A range of techniques were used in order to compare whether two pieces of MIDI sounds are similar - including combining melody extraction, standardization and similarity functions to improve the results.

Sound archival is also part of the problem, as mentioned in the introduction [Gorder 2006]. This involves issues of data structures, but also recording formats and compression. Data heterogeneity also plays a major role, a common problem in eScience. The recording is just the tip of the iceberg. It is the result of the scientific methodology used to decide what to record (and when, and where, and with which device). Hence, designing animal sound management platforms is a hard task, especially if different sound collections are to be somehow integrated.

Management of animal sounds requires cooperation from domain experts - the biologists. Different animal groups present distinct forms of sound production and, therefore, different sounds are produced. In general, these sounds evolve under the pressure of sexual selection, promoting significant differences between different species. As a result, almost each animal species in the world produces a distinct sound. However, there is a complication: there are multiple levels of complexity in sound production. For example, crickets and frogs generally present simple advertisement calls (i.e. just one type of note) with well conserved structure (even between individuals of different populations). On the other hand, birds and mammals may present different notes (i.e. complex calls) in the same call, may present differences between populations (i.e. the calls from individuals of the same species from the northern range may be different from that from its southern distribution), between sex (females *vs.* males), age (juveniles *vs.* adults), social context (e.g., isolated individuals *vs.* grouped individuals; or territorial context *vs.* reproductive context). The social context may also influence sound emission of frogs and crickets [Gerhardt and Huber 2002] [Wells 2007].

As such, standard similarity techniques may not suffice – queries have to consider additional factors, such as geographic location. Also, it is possible that specific algorithms need to be developed for distinct taxonomic families. Indexing is also complicated by such particularities – multidimensional structures might be adequate, the problem is to select the appropriate features to index. Spectrograms can be an important resource to help researchers visualize animal sound. Another interesting challenge related to acoustic data is the automatic creation of metadata, to help reinterpretation and re-analysis of the data [Wimmer et al. 2010].

If users know some metadata (e.g., taxon, location) then a possible approach is to perform the query in multiple steps, starting with textual filtering. However, even text-based queries also pose open problems – e.g., [Daltio and Medeiros 2008]. Not only does metadata content has to be individually curated, for each recording, but there are other domain-specific concerns, such as species naming (and distinct versions of taxonomic classifications), ambiguity in location identification, and diversity imposed by the variety of collection methodologies, affecting data and metadata.

One must also consider sound management software. [Barrington et al. 2007] have proposed a system for audio information retrieval using semantic similarity . The system retrieves generic sound records through ranking items in a database using semantic similarity, rather than acoustic similarity. In order to improve the results, queries are performed using both acoustic algorithms and semantic metadata queries.

In the past, biologists did not always have access to devices to analyze sounds. With the increase of computational power of personal computers, many softwares for acoustic analysis were developed. One example is the research performed by [Wimmer et al. 2010], a workbench that helps ecologists to automatically recognize animal sounds (ASR - Automatic Speech Recognition), annotate and remove sound noise from sounds that can be saved in a centralized database. [Bardeli 2009] created a method for similarity search in animal sound databases. Bardeli also proposed algorithms for feature extraction, indexing and retrieval for animal sound databases. Similarly, [Gunasekaran and Revathy 2010] proposed an algorithm for classification and retrieval of animal sounds. The algorithm is based on fractal dimensions analysis to compute k-nearest neighbors. Such systems, however, are not generic and have their limitations – e.g., the fact of proposing algorithms for specific groups, usually does not cover important features that are essential in groups from similar domains. Then, there are lack of systems that are focused on general domains, covering the necessities of different groups.

A final direction concerns the so-called citizen science (in which citizens contribute to scientific efforts by recording observations, or helping annotate data [Dickinson et al. 2010]). This has given rise to several major data management projects in which citizens provide lots of primary and secondary data - with consequent research on cleaning, indexing and processing such data. Examples are bird-watcher citizen science sites (e.g., the projects in the Cornell Ornithology Laboratory[1]) or the FrogWatch project [2] in which people learn to recognize frog calls, and then report hearing them in nature, to find out about wetland ecology. Most of these projects involve entering text, sometimes pictures, but to the best of our knowledge there is no project involving sound management, or sound mining. Some sites, like FrogWatch, publish frog call sounds so that volunteers may learn how to recognize them in nature. The number and variety of possibilities is limited, and there is no linkage to more advanced data management tools.

Analyzing these research efforts, it is clear that animal sound recordings need specific technologies to treat them. For instance, traditional data mining techniques are not able to be employed for these data. Then, although the research cited in this section solves specific issues, all of them have common challenges: how to manage large amounts of heterogeneous animal sound data.

3. Managing Large Amounts of Sounds - A Case Study

Between 1970 and 2010, Prof. Jacques Vielliard *(in memoriam)* and students recorded about 20.000 sounds, mainly birds in South America, mostly recorded in magnetic tapes – now the Fonoteca Neotropical "Jacques Vielliard", UNICAMP. These recordings are undergoing digital conversion. Although this conversion started many years ago, due to its complexity, about 12.000 sounds from the collection have been saved in digital format so far.

This collection - the largest of its kind in the Neotropics - is frequently visited by biologists for their research. It does not support online access, which hampers its widespread use. Thus our first step was to develop a web system to publish sound record metadata, and to allow scientists to upload and download recordings. Our ultimate goal is to increasingly enhance the functions offered at this site, to support novel query and access operations.

Sound management systems, however, are not new - e.g. see Motorola/Shazam Music Recognition Software[3]. What is then so different about management of animal recordings? First, they are often made under difficult conditions, presenting lots of background noise. Also, when recorded in their natural habitats, many animal sounds appear in a single recording, there being a need to identify individuals (or at least individual species). Even if one assumes recordings are performed under ideal conditions, such sounds are very much season, geographic distribution, and social context sensitive (see section 2).

The design of our web system is not a simple task. Biologists primarily use spreadsheets in order to help them manage the collection of sounds, mostly recording metadata. However, these are not adequate for more advanced resources to manage the sound records. Using spreadsheet helps them manage the sounds' metadata, but do not help the management of sound recording, or advanced query processing.

Here, interoperability is a major problem, if one wants to give access to several sound repositories. There arise many issues such as the definition of metadata standards

[1] http://www.birds.cornell.edu/citsci/projects

[2] http://www.aza.org/frogwatch/

[3] http://www.shazam.com/music/web/pages/shazamid.html

or common query vocabularies (and thus ontologies). At present, most recordings are exchanged via email or even by the post office. Second, sound analysis is often based on statistical studies, and queries for records are based on metadata. Query by content is limited, due to the particular characteristics presented by distinct animal groups - see section 2.

4. Publishing Animal Sounds on the Web

This section describes our prototype available at http://proj.lis.ic.unicamp.br/BioProj. It was developed as a web environment to help the management of the sound collection at Fonoteca Neotropical. Its long term goal, however, is to support two kinds of users and usages. First, it is intended to help biologists around the world who want to upload and share bird sounds recorded by them. Thus, it aims to support cooperative research projects that involve animal recording analysis - e.g. in biodiversity, behavior and ecology. Second, it will also contribute to development of citizen science initiatives [Dickinson et al. 2010], allowing non-experts to contribute to enhance data on sound collections, and to learn more about species. The prototype was developed in Java and JSF (Java Server Faces). It was also used Hibernate framework (to perform data persistence), Ajax (Rich Faces Framework) and MySQL DBMS.

Currently, the prototype has three main functionalities:

- Navigate the whole collection of sounds
- Perform metadata-based queries
- Upload bird sounds

4.1. Navigate the whole collection of bird sounds

This functionality allows users to see all recordings, ordered by scientific taxonomy. Though simple, this kind of functionality is a first step towards a more sophisticated navigation system, whose specification requires user feedback. Figure 1 shows the main screen, where biologists are able to view taxonomy data associated with each recording. For instance, record with identification 7 concerns a bird song – Phylum Chordata, Class Aves, and so on. Users can also download non-copyrighted sound files, and view recording metadata by clicking on the ”details” button. Figure 2 shows extra metadata of a selected sound record - e.g., information such as the time and date that the sound was recorded, country, temperature of the air at the moment of the recording. Metadata is extracted from the Excel files.

4.2. Metadata-based textual queries

This functionality allows biologists to retrieve only the desired sound record. Figure 3 shows part of the query form, with multiple textual filters, including specific taxonomy classification terms, country or city where the sound was recorded. For example, Figure 3 shows a query that retrieves all recordings that were recorded in Brazil, whose Phylum is *Chordata*, Class *Aves* and Order *Tinamiformes*. Users can subsequently retrieve more information and download sounds – see Figure 1.

Figure 1. Page to navigate the whole collection of bird sounds

Figure 2. Form containing details about one sound record metadata

4.3. Upload sounds

This functionality allows biologists from Fonoteca Neotropical to upload sounds. Another goal is to allow biologists around the world to upload and share bird sounds recorded by them. The deposit of recordings must be mediated by expert curators. Figure 4 shows the form to upload sounds. It may eventually be used to support citizen science, e.g. people who like to record animal sounds as a hobby will be also able to publish and share their recordings. In this case, submitted data must be differentiated from curated scientific data, posing challenges in distinct management functions.

4.4. Rationale Behind the Prototype

In this project, we have adopted a collaboration policy that has proved to be successful in previous multidisciplinary projects of ours [4]. Instead of designing and developing a complex system with many query possibilities (e.g., sound mining, or indexing), we developed a first prototype to help scientists manage their data. Once they test and approve a prototype, then we can improve on it, on a continuous evolution lifecycle, rapid prototyping style. This has the advantage of ensuring fast feedback and in attracting a larger number of users. We point out that, even if relatively simple, this kind of web system is very much in demand by experts and, in our case, it has the advantage of the richness of the underlying data collection.

[4] http://www.lis.ic.unicamp.br/projects/biocore

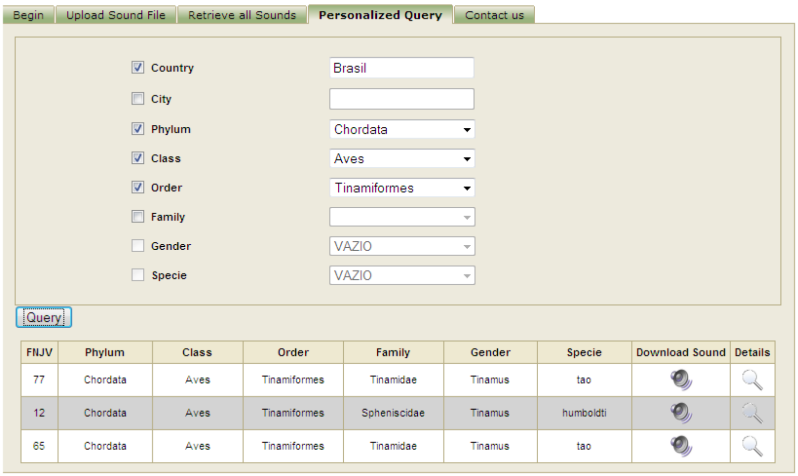

Figure 3. Form to perform metadata-based queries

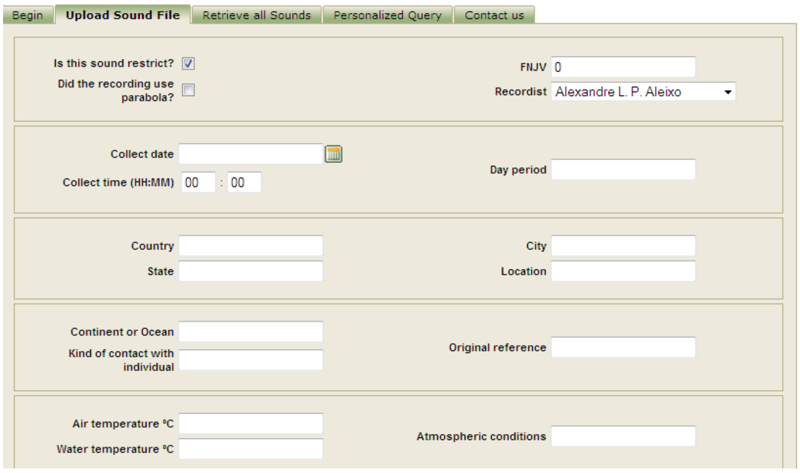

Figure 4. Page to upload bird sounds

5. Final Considerations and Ongoing Work

This research is being motivated by challenges faced by Fonoteca Neotropical on managing large amounts of animal recordings. The current collection, which has more than 1 terabyte of digital sound data, is growing up. Only half the recordings have been digitalized. We are now working on publishing all metadata online, linked to the recordings.

Ongoing and future work involve attacking some of the challenges mentioned, with the ultimate goal of designing and developing a full-fledged animal sound exploratory system for biologists and citizen science. One issue is to make copyright sounds available for biologists that want to use them for research. To do this, it will be necessary to provide distinct access control mechanisms. This control may be also exercised for all downloads, and will create new data files that can be further used for data mining, for example scientists can find other scientists that are working with the same species calls, improving interaction among scientific community and improving worldwide science.

6. Acknowledgments

We thank CAPES, CNPQ and FAPESP for financial support. We also thank biologists from Fonoteca Neotropical ”Jacques Vielliard”, who provide us support in this research.

References

Bardeli, R. (2009). Similarity search in animal sound databases. *Multimedia, IEEE Transactions on*, 11(1):68–76.

Barrington, L., Chan, A., Turnbull, D., and Lanckriet, G. (2007). Audio information retrieval using semantic similarity. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE Int. Conf. on*, volume 2, pages II–725 –II–728.

Cobos, M. and Lopez, J. (2010). Listen up - the present and future of audio signal processing. *Potentials, IEEE*, 29(4):40 –44.

Daltio, J. and Medeiros, C. (2008). Aondê: An ontology web service for interoperability across biodiversity applications. *Information Systems*, 33(7-8):724–753.

Dickinson, J., Zuckerberg, B., and Bonter, D. (2010). Citizen Science as an Ecological Research Tool: Challenges and Benefits. *Annual Review of Ecology, Evolution, and Systematics*, 41(1).

Gerhardt, H. and Huber, F. (2002). *Acoustic communication in insects and anurans: common problems and diverse solutions*. University of Chicago Press.

Gorder, P. (2006). Not just for the birds: archiving massive data sets. *Computing in Science Engineering*, 8(3):3 –7.

Gunasekaran, S. and Revathy, K. (2010). Content-based classification and retrieval of wild animal sounds using feature selection algorithm. *Machine Learning and Computing, Int. Conf. on*, 0:272–275.

Uitdenbogerd, A. and Zobel, J. (1999). Melodic matching techniques for large music databases. In *Proc. of the 7th ACM Int. Conf. on Multimedia (Part 1)*, MULTIMEDIA ’99, pages 57–66, New York, NY, USA. ACM.

Weihs, C., Ligges, U., Morchen, F., and Mullensiefen, D. (2007). Classification in music research. *Advances in Data Analysis and Classification*, 1(3):255–291.

Wells, K. (2007). *The ecology & behavior of amphibians*. University of Chicago Press.

Wimmer, J., Towsey, M., Planitz, B., Roe, P., and Williamson, I. (2010). Scaling Acoustic Data Analysis through Collaboration and Automation. In *2010 IEEE 6th Int. Conf. on e-Science*, pages 308–315. IEEE.