Ecological Informatics 12 (2012) 50–67

Contents lists available at SciVerse ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology

Eric P. Kasten [a,*], Stuart H. Gage [b], Jordan Fox [c], Wooyeong Joo [d]

[a] Clinical and Translational Sciences Institute, Biomedical Research Informatics Core, Michigan State University, 888 Wilson Road, Room 100, East Lansing, MI 48824, USA
[b] Department of Entomology, Michigan State University, 1405 S. Harrison Road, Suite 101, East Lansing, MI 48823, USA
[c] Global Observatory for Ecosystem Services, Michigan State University, 1405 S. Harrison Road, Suite 101, East Lansing, MI 48823, USA
[d] Remote Environmental Assessment Laboratory Michigan State University, 1405 S. Harrison Road, Suite 101, East Lansing, MI 48823, USA

ARTICLE INFO

Article history:
Received 15 June 2012
Accepted 27 August 2012
Available online 10 September 2012

Keywords:
Acoustics
Sound archive
Normalized difference soundscape index
Soundscape ecology
Symbolic representation
Unsupervised learning

ABSTRACT

Acoustic signals constitute a source of information that can be used to measure the spatial and temporal distributions of vocal organisms in ecosystems. Measuring and tracking those species that produce sounds can reveal important information about the environment. Acoustic signals have been used for many years to census vocal organisms. Moreover, acoustics can be used to compute indexes for measuring biodiversity and the level of anthropogenic disturbance. We developed the software and system that automate the process of cataloging acoustic sensor observations into the Remote Environmental Assessment Laboratory (REAL) digital library that can be accessed through a website (http://lib.real.msu.edu). The REAL digital library enables access and analysis of collected acoustic sensor observations. We report on current library status and the mechanisms that enable the selection, extraction and analysis of acoustic data to support investigations on automating species census as well as measuring diversity and disturbance. We implemented numeric and symbolic search mechanisms and unsupervised learning techniques to ease retrieval of acoustic information, including recordings and processed data, pertinent to visitor goals.

1. Introduction

With the growth and expansion of human populations has come an increasingly greater need to understand the dynamics of ecosystems and their complex interactions (Michener et al., 2001). As urban areas continue to expand, fragile surrounding ecosystems are often affected as they are redesigned to support urban infrastructure. In Michigan, urban populations are estimated to increase 180% by the year 2040 (Skole et al., 2002). When human development impacts critical components or linkages within an ecosystem, the function and structure of that ecosystem are dramatically altered (Vitousek et al., 1997). The need to better understand these effects has led to the development of ecological indicators, which are variables measured within an ecosystem that contain information regarding the degree of ecosystem stress and perturbation. Vitousek et al. (1997) argue that these ecological indicators need to capture the complexities of the ecosystem, yet they need to be routinely and easily monitored.

Acoustics as an ecological attribute has the potential to increase our understanding of ecosystem change due to human disturbance, as well as provides a measure of biological diversity and its subsequent change over time (Joo et al., 2011; Sueur et al., 2008; Truax, 1984; Wrightson, 2000). The analysis of entire soundscapes may also produce valuable information on the dynamics of interactions between ecological systems in heterogeneous landscapes (Carles et al., 1999; Joo et al., 2011; Pijanowski et al., 2011b). Moreover, timely analysis and processing enables rapid delivery of important environmental information to those responsible for conservation and management of our natural resources, and can promote public involvement through public access to ready information about the environments in which we live. For instance, increased interest in renewable energy sources has driven the development of wind resource areas and the need to better understand the unintended impact of wind farms on wildlife. In turn, state and federal agencies have put forth guidelines for evaluating the potential effects that a wind farm might have on wildlife that includes acoustic monitoring (Anderson et al., 1999; Michigan Department of Labor and Economic Growth, 2005; United States Fish and Wildlife Service, 2003). In addition, the National Park Service has supported studies to measure the temporal variability of the soundscape in US National Parks (Krause et al., 2011).

Acoustic signals have been used for many years to census vocal organisms. For example, the North American Breeding Bird Survey, one of the largest long-term, national-scale avian monitor programs, has been conducted for more than 30 years using human auditory and visual cues (Bystrak, 1981). The North American Amphibian Monitoring Program is based on identifying amphibian species primarily by listening for their calls (Weir and Mossman, 2005). Recent

* Corresponding author.
E-mail address: joowooye@msu.edu (W. Joo), kasten@msu.edu, http://www.msu.edu/~kasten (E.P. Kasten), gages@msu.edu, http://www.real.msu.edu/~gages (S.H. Gage), foxjo@msu.edu (J. Fox).

1574-9541/$ – see front matter
http://dx.doi.org/10.1016/j.ecoinf.2012.08.001

advances in sensor networks enable large-scale and automated collection of acoustic signals in natural areas (Estrin et al., 2003). The systematic and synchronous collection of acoustic samples at multiple locations, combined with measurements of ancillary data such as light, temperature, and humidity, can produce an enormous volume of ecologically relevant data. Transmuting this raw data into useful knowledge requires timely and effective processing and analysis.

The main contribution of this paper is to describe the current status of the Remote Environmental Assessment Laboratory's (REAL) digital library of acoustic sensor observations collected using in-field acoustic sensor units. This library enables search results to be heard, visualized, plotted and downloaded to construct data sets using geographic, temporal, symbolic and data cluster-based search criteria for retrieving sets of acoustic observations. We investigate symbolic and unsupervised learning techniques that aim to ease the extraction of data pertinent to library visitor needs. Unsupervised learning refers to techniques that group observations according to some metric of similarity (e.g., Euclidean distance) without a priori categorical knowledge (Duda et al., 2001). This process is often called data clustering because data are clustered into "natural groupings." We describe how these techniques are used to support query-by-example library searches.

The remainder of this paper is organized as follows. Section 2 describes our motivation for constructing the REAL digital library. Section 3 presents a description of the current status of the library and describes the example study area that we use throughout this article. Basic search operations are discussed in Section 4. Symbolic representations and unsupervised learning to support query-by-example search operations are described in Sections 5 and 6, respectively. Section 7 describes our evaluation methods and Section 8 presents our experimental results. Finally, we conclude and describe future work in Section 9.

2. Motivation for the REAL digital library

Acoustic measurement exhibits several desirable properties. First, whereas traditional approaches to species census and measuring biological complexity are labor intensive, the infrastructure used to collect acoustic data can be managed by non-expert field staff (Hobson et al., 2002; Michener et al., 2001; Thompson et al., 2001; West et al., 2001). Second, continuous and stationary acoustic monitoring reveals spatio-temporal patterns that cannot be captured in site-by-site observations. By remaining in one location and monitoring continuously, acoustic information can reveal changes in ecosystems over diurnal, monthly, seasonal, yearly, and other temporal scales (Truax, 1984). Third, because acoustic monitoring systems can simultaneously monitor in multiple locations, acoustical variances can be compared to environmental heterogeneity and landscape structure (Michener et al., 2001; Pijanowski et al., 2011b; Thompson et al., 2001; West et al., 2001). Fourth, microphones can collect data from all directions simultaneously despite occlusions, such as trees or buildings, and cover of darkness. Finally, ecological acoustics can be measured automatically with minimal human interference. Recording technology operates independently in the field, thereby allowing observation to take place without interference generated by human presence (West et al., 2001). However, when ecological sensor platforms collect data on a continuous or regular periodic basis, the sheer volume of the data might preclude the extraction of pertinent information of interest. Addressing these problems will likely become increasingly important in the future as technology improves and more sensor platforms and sensor networks are deployed (Pijanowski et al., 2011a; The, 2020 Science Group, June/July 2005). The REAL digital library represents a step toward applying automated processing to organize large archives of acoustic sensor data with the goal of providing novel data representations and search mechanisms that ease the retrieval of pertinent information. Below we suggest two potential application challenges that automated sensor data acquisition and cataloging in archives, such as the REAL digital library, can help address.

2.1. Automated census

Traditionally, one of the most commonly used survey methods for identifying bird species and estimating abundance has been the point count survey (Ralph et al., 1995; Rosenstock et al., 2002; Thompson, 2002). The point-count method depends on a human observer to identify birds using acoustic and visual cues within a fixed distance at specific points in an area block or along a line transect during the breeding season (Hutto et al., 1986; Ralph et al., 1995). Several studies have expressed concerns about using the point count method to survey birds because of inconsistencies in detection probability among species and across habitats or over time (Rosenstock et al., 2002; Thompson et al., 2001). They also noted that even highly skilled observers cannot detect every bird that is present or singing simultaneously. Others have raised concerns regarding the high variability found among observers with respect to their ability to accurately conduct surveys (i.e., experience and survey skills, age, and hearing loss) (O'Connor et al., 2000). Point-count results can vary considerably due to diurnal and seasonal changes that occur during survey periods (Canadian Wildlife Service, 2006). In addition, when using only the point-count method, there is no validation of the species identified by an observer. The traditional point count method, although valuable, has several deficiencies that can be improved through the use of automated techniques for capturing, organizing and searching acoustic observations.

2.2. Studying soundscape ecology

Ecosystem sounds create a soundscape, comprised of acoustic periodicities and frequencies emitted from the ecosystems' biophysical entities (Qi et al., 2008; Schafer, 1977; Truax, 1978, 1999). Soundscape ecology is "the study of systematic relationships between humans, organisms, and their sonic environment" (Pijanowski et al., 2011a, 2011b; Schafer, 1994) or "the study of the effects of soundscape on the physical responses or behavioral characteristics of living organisms in the system" (Truax, 1999). Additionally, some studies explored the relationship between sounds and the images of their associated locations and investigated the influence of sounds on landscape value and characteristics (Adams et al., 2006; Carles et al., 1999), because visual sensing of landscapes by humans and organisms is closely correlated with the sound experience (Southworth, 1969).

The use of environmental acoustic sensor data has rarely been considered as informative or as a useful quantitative measure in the environmental sciences (Carles et al., 1999). On the other hand, it has been shown that modification of natural systems affects biodiversity, ecosystem processes and functions (Vitousek et al., 1997), and the associated soundscape (Porter et al., 2005). In some cases, the characteristic soundscape has even disappeared from the system after modification (Barber et al., 2009; Joo, 2008; Warren et al., 2006). Moreover, the processing and analysis of complex acoustic data requires advanced computation technologies and newly developed algorithms. Recently, it has been realized that analyzing the structure and patterns of the soundscape can provide useful spatiotemporal information (Gage et al., 2001; Joo, 2008; Joo et al., 2011; Qi et al., 2008; Schafer, 1977). Notably, advances in modern technology and high-powered computation instruments enable novel approaches for processing and quantifying environmental acoustic data (Estrin et al., 2003; Gage et al., 2001; Joo, 2008; Kasten et al., 2007, 2010; Qi et al., 2008) and extracting a variety of ecological information such as vocalizing species identification (Chou et al., 2007; Kasten et al., 2010), diversity metrics (Sueur et al., 2008), and the effects of human noise on natural and human systems (Hannah et al., 1994; Krausman et al., 1986; Neumann and Merriam, 1972; Romano et al., 2004).

By advancing our capacity for collecting, organizing and searching large archives of acoustic data, we strive to enable the study of ecological objectives that fall under the research themes proposed by Pijanowski et al. (2011b). These research themes include questions related to: (1)improving the measurement and quantification of sounds, (2) improving our understanding of spatial–temporal dynamics across different scales and how environmental covariates impact sound, (3) assessing the impact of the soundscape on humans and wildlife, and (4) assessing the impact of human activity on soundscapes.

3. Background

In this section, we first discuss other projects and acoustic libraries of natural sounds that are related to our work. Second, we present an overview on the collection of in-field acoustic observations and the process by which they are cataloged in the REAL digital acoustic library. Third, we discuss the current status and contents of the library. Finally, we describe the example study area that we use in this article to illustrate our approaches for search and retrieval of acoustic sensor data.

3.1. Related projects and libraries

There are several other libraries of environmental and natural acoustics (Cornell Laboratory of Ornithology, 2009; Department of Evolution, Ecology and Organismal Biology, 2009; Florida Museum of Natural History, 2009; The British Library, 2009). Typically, these libraries are organized by criteria such as location, taxonomy and common name. Although these databases provide important reference to scientific information and bioacoustics they do not address the need for automated organization and searching of acoustic sensor data. When data is continuously collected, automated processing facilitates the organization and searching of the resulting data repositories. Without timely processing, the sheer volume of the data might preclude the extraction of information of interest. Addressing these problems will likely become increasingly important in the future as technology improves and more sensor platforms and sensor networks are deployed.

(Mellinger and Clark, 2006) created Mobysound to provide public access to a repository of data sets to support research in automatic recognition of marine animal vocalizations. This database comprises a set of recordings that can be used by researchers for training neural networks and other automated call recognition systems. Such publicly available data sets provide an important service to the research community by providing common access to data sets that can be used for comparing different recognition methods. Whereas the REAL digital library targets the organization and searching of acoustic recordings to ease visitor access to sensor data, Mobysound enables topic and purpose specific data sets to be downloaded to support the study of automated recognition of marine mammal calls. It is anticipated that subsets of REAL library data will enable the construction of new data sets that can be used to further research on call recognition.

Classification of organisms based on their vocalizations is an active area of research (Berndt and Clifford, 1994; Fagerlund, 2007; Fagerlund and Härmä, 2005; Kogan and Margoliash, 1998; Somervuo and Härmä, 2004; Vilches et al., 2006). For instance, Mellinger and Clark (2000) addressed classification of whale songs, with specific application to identification of bowhead song end notes, using spectrogram correlation, Chou et al. (2007) used HMMs for recognition of bird species based on song syllables, and Kasten et al. (2010) used anomaly detection for detection and extraction of acoustic events for classification of bird species. Moreover, researchers have also used computer vision techniques to help in the recognition of recorded music. For example, Ke et al. (2005) used such techniques to produce signatures from acoustic spectrograms for recorded songs. Automated recognition and classification of bioacoustic signals will likely be important for cataloging and searching acoustic sensor data. However, the techniques that we describe in this article use constant space representations and unsupervised learning to organize and search entire recordings based on their similarity. Unlike classification methods that target specific sets of vocalizations, our techniques enable summarization and organization of acoustic clips without knowledge of a priori classification categories.

3.2. Data collection overview

Automated collection of acoustic and ancillary data is an important goal of this project. Sensor clusters comprise two or more sensor platforms and an optional cluster server. Each sensor platform collects data at pre-programmed times with minimal site disturbance or regular human intervention. A typical platform comprises a pole-mounted sensor unit and optionally a solar panel coupled with a deep cycle battery for providing power over extended periods. Platforms include several custom sensor unit types, designed and implemented by REAL personnel, and the SongMeter™ produced by Wildlife Acoustics® (Wildlife Acoustics, 2009). The custom sensor units were designed to record acoustic clips and transmit them over a wired or wireless network directly to the sensor data depot or to a regional cluster server for later relay to REAL to the data depot. When using SongMeters™, data must be manually collected and later uploaded to the REAL library system for batch processing.

3.2.1. The depot

The sensor data depot provides near real-time processing and analysis of sensor data as it is transmitted from sensor units and cluster servers, enabling early vetting of sensor unit operation and data collection. Once a user logs into the system and selects a project, the sensors associated with the project are accessible. The user can access specific sensor observations based on the time and location of the observation. The depot enables the acoustic observation to be viewed (oscillogram, spectrogram, frequency bins) and listened to.

3.2.2. Digital library

After sensor data arrives in the depot, it is subsequently processed for cataloging in the REAL digital library. We implemented a relational database schema to enable rapid access to large sensor data sets. As shown in Figs. 1 and 2, library visitors can access sensor data collections through a browse or search interface, and can download data to a local computer for customized analysis, or use the general analysis routines provided through the library's web-based workbench. Our data acquisition system has the capacity to automatically populate this acoustic library at our sensor observatory as data is transmitted from in-field deployments.

3.3. Digital library components and processes

The digital library was designed and implemented to accommodate a large number of acoustic recordings and to support the storage, processing, and analysis of these data. There are seven basic components involved in library data processing including: collection, upload, data processing, archiving, analysis, access, and interpretation. Each of these components are described below and related to the diagram shown in Fig. 4 as follows. Collected recordings are uploaded to temporary storage (1). If the acoustic sensors are online and can transmit data to the library, the recordings from the sensors are sent directly to the depot. Otherwise, recordings are manually retrieved from the sensors and subsequently uploaded via the upload toolbox (2). The recordings are processed by the sound processing module (3). Once the recordings have been processed, the raw recordings and processed data are loaded into the library database by the database integration module (4). Raw soundscape recordings, metadata, and the results of initial computational processing are linked using relational database technology. Soundscape analysts and users access this information using the acoustic library access and output modules (5). The access module enables browsing, query, search and cluster retrieval using relational queries. These queries are focused through selection of a project, geographical location, date and time. The visual, auditory and digital results of searches are provided to the user, and query results can be downloaded using the download toolbox (6). Desktop analysis of the sensor recordings can be conducted using other statistical or signal processing software, enabling user specific analysis and interpretation of the soundscape (7).

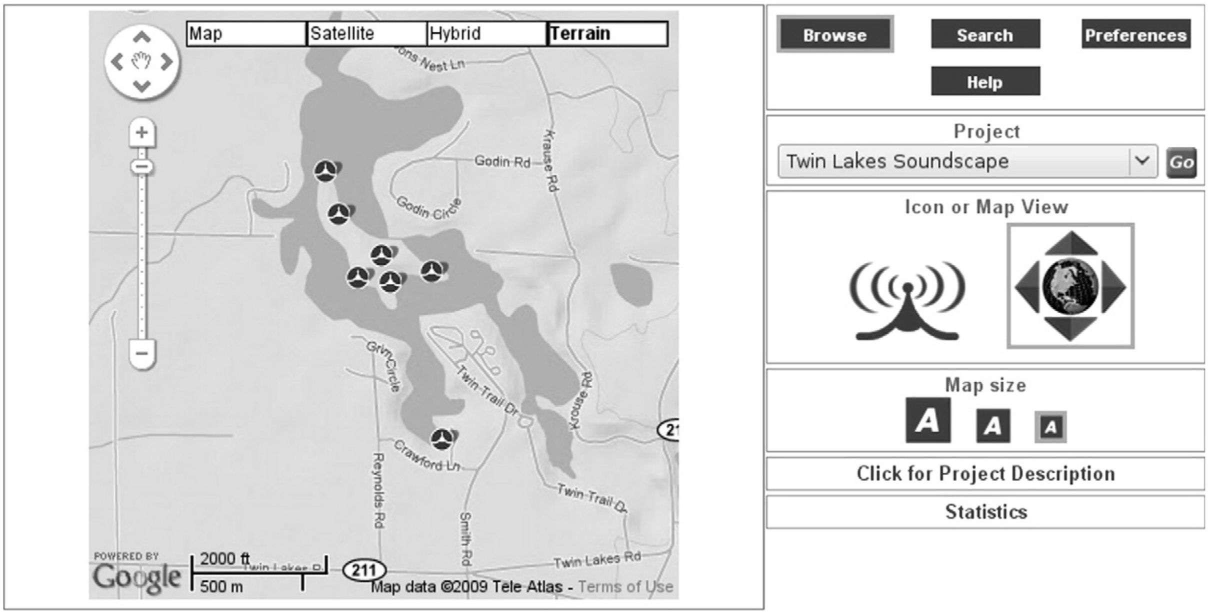

Fig. 1. Library browse screen for the Twin Lakes Soundscape, Cheboygan, MI. When browsing, a library visitor can access data for a specific sensor unit by clicking on the geographic location of the sensor unit as indicated through a Google™ enabled geographic interface.

3.4. Library catalog

The real library can be accessed through a web site at http://lib.real.msu.edu. This web site is supported by several technologies including: the Apache web server (The Apache Software Foundation, 2009), JavaScript (Flanagan, 2006) a MySQL database server (Sun Microsystems, 2009), the PHP scripting language (Achour et al., 2009), and the Dynamic River distributed stream processing toolbox (Kasten et

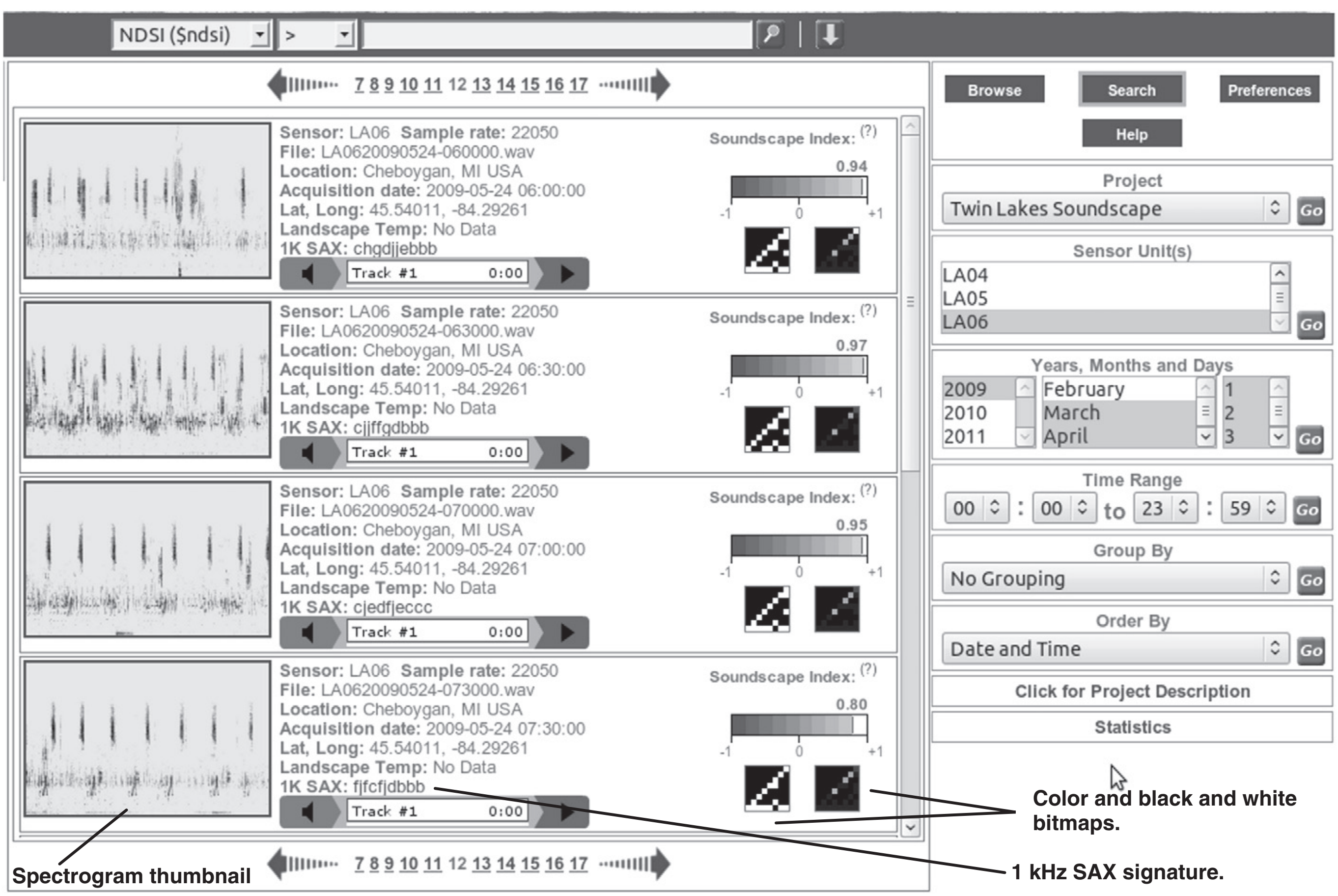

Fig. 2. Library search screen. Each summary box displays a spectrogram thumbnail, basic temporal and location information, a normalized difference soundscape index (NDSI), 1 kHz Symbolic Aggregate approXimation (SAX) string, and color and black-and-white SAX bitmaps. In addition to the time, date and location search interface on the right, the library can be searched using symbolic and cluster-based mechanisms.

Fig. 3. Example study area. Twin Lakes, Cheboygan, Michigan. Insets from left to right, sandhill crane (*Grus canadensis*), Canadian geese (*Branta canadensis*), common loon (*Gavia immer*), American red squirrel (*Tamiasciurus hudsonicus*), double crested cormorant (*Phalacrocorax auritus*).

al., 2007). Currently, the REAL digital library comprises 20 projects that contain more than 2.8 TB of digital acoustic data recorded in more than 1 million files that can be accessed online, and continues to grow with the addition of new projects and the upload of sensor data. Projects cover time frames that span a few days to several years, including one project that spans the period from 2001 until the present. Locations include among others: Twin Lakes, Cheboygan, MI; The Kellogg Biological Station Long Term Ecological Research Station, Hickory Corners, MI; Crane Island, Montmagny, Quebec, Canada; Santa Margarita Ecological Reserve, San Diego, California; the Samford Ecological Research Facility, Queensland, Australia; the University of Notre Dame's Environmental Research Center, Land O'Lakes, Wisconsin, and the University of Wisconsin's Trout Lake Station, Boulder Junction, Wisconsin. Typically, the library comprises 30 or 60 s sound clips collected every half-hour at each location at a rate of 22,050 samples/s, sufficient for capturing the vocalizations of most organisms in the Great Lakes, temperate region. However, the library can support higher sampling rates and longer durations if required. Acoustic project goals vary from species census to comparison of urban/rural soundscapes to ecological education. Descriptions of past and current REAL projects can be found at http://www.real.msu.edu/projects/.

3.5. Example study area: Twin Lakes

Twin Lakes is located in Grant Township, Cheboygan County, Michigan. As shown in Fig. 1, Twin Lakes is a chain of seven small interconnected basins separated by narrow channels. The surface area is only 207 acres but the shoreline length is 46,262 ft due to basin configuration. Twin Lakes is located in forested land comprising mixed coniferous and deciduous trees with predominately sandy soil. The lake chain has a few small creeks as inlets and has only one outlet, Owens Creek. The lakes are primarily groundwater fed. Most basins are 25–45 ft deep with one basin extending to 73 ft in depth. One of the key features is an uninhabited island, located near the center of the lake basins. The island is co-owned by the State of Michigan and the Federal Government.

Amphibians such as green frog, spring peeper, and leopard frog inhabit the aquatic vegetation. As shown in Fig. 3, many species of birds, amphibians and mammals inhabit the ecosystem including bald eagles, osprey, Caspian terns, belted kingfishers, blue herons, trumpeter swans, common loons, mergansers, pheasant, ruffed grouse, wild turkey and many smaller woodland birds. Field and forest dwellers include black bear, white-tailed deer, coyote, fox, beaver, river otter, marten, raccoon, rabbit, squirrel and chipmunk.

Twin Lakes include a significant population of human inhabitants. Gannon (1974) noted that there were approximately 50 widely scattered, mostly summer cottages on the Twin Lakes shoreline, and that this lake chain has the highest shoreline development factor of any inland water body in Cheboygan County. Because of the configuration of these lakes, there is more shoreline in relation to the size of the lakes and is subject to much greater potential for development and recreational pressure. Since 1974 the number of dwellings on these lakes has more than doubled, and many of the dwellings are now occupied by permanent residents.

Wildlife Acoustics® SongMeters™ were used to record acoustic observations at this site. Recordings were made every half hour for a duration of 60 s at a sampling rate of 22,050 samples/s, providing a usable frequency range up to 11 kHz. As shown in Figs. 1 and 6 sensor units were located on the island and 3 units on the mainland. Data was collected manually for later cataloging in the library. There are currently more than 205,582 audio recordings, using more than 515 GB of disk storage, cataloged for the Twin Lakes Soundscape in the REAL digital library.

4. Organization and searching

In this section, we first describe how the library can be searched by location, date and time. Then, we introduce a normalized difference soundscape index (NDSI) and discuss how observations can be retrieved using basic frequency and the NDSI.

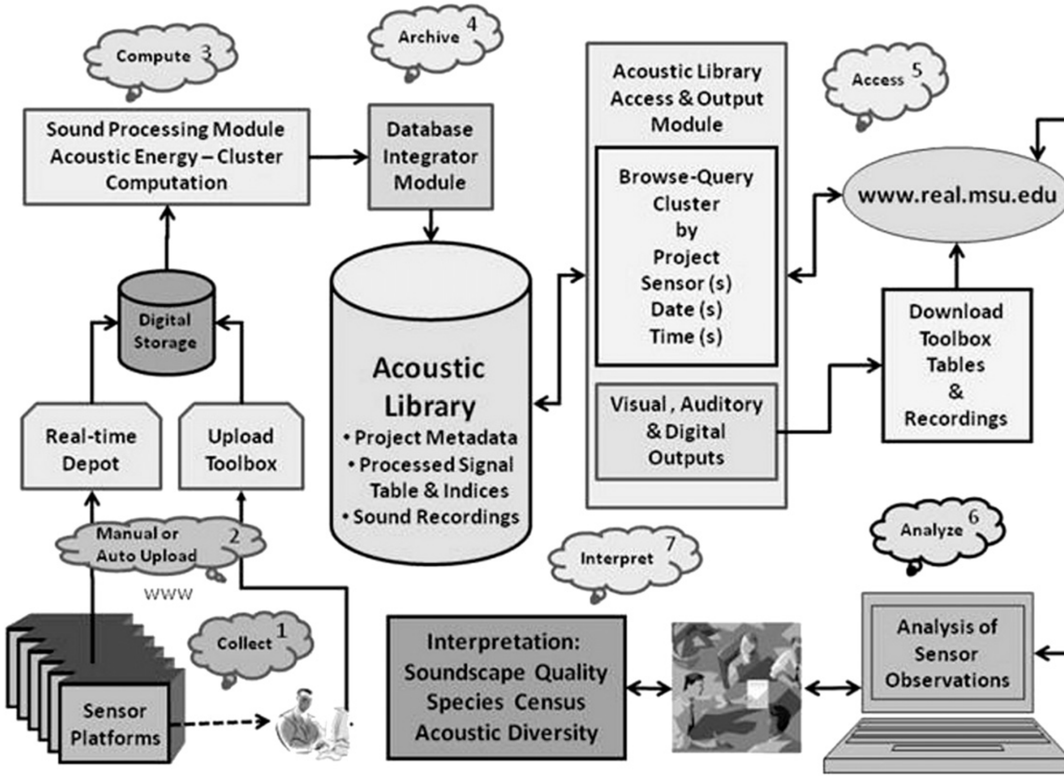

Fig. 4. Library components and processes. A diagram of the acoustic library showing the processes of: soundscape data collection (1), data uploading (2), automated computing of acoustic energy and clustering (3), archiving computed results and soundscape recordings into the acoustic library (4), access via relational database to browse, query or cluster recordings in the acoustic library and output generation (5), desktop analysis of downloaded query results (6) and interpretation of recordings, patterns and analysis (7).

4.1. Searching by location, date and time

As shown in Fig. 2, a library visitor can select acoustic observations by project, sensor unit, date and time using the panel on the right. In addition, results can be grouped and ordered by date, time or sensor unit. Results are shown as spectrogram thumb nails with basic information such as sensor unit identifier, location and collection date and time. A spectrogram depicts frequency on the vertical axis and time on the horizontal axis. Shading indicates the intensity (power spectral density) of the signal at a particular frequency and time. We computed spectrogram power spectral density using Welch's method (Welch, 1967) with a 50% overlap between segments of length 1024 that were filtered using a Hamming window prior to processing with the fast Fourier transform. Each observation can be listened to using the embedded audio player provided. Other visual and symbolic representations are also shown including: a bar plot of the NDSI (see Section 2), a 1 kHz SAX string (see Section 5), and a pair of SAX bitmaps (see Section 6). A larger, more detailed spectrogram, oscillogram and plot of the 1 kHz frequency bins for a specific observation can be viewed by clicking on the thumb nail.

4.2. Searching the frequency space

4.2.1. Normalized difference soundscape index (NDSI)

The goal of our normalized difference soundscape index (NDSI) is to estimate the level of anthropogenic disturbance on the soundscape by computing the ratio of human-generated (anthrophony) to biological (biophony) acoustic components found in field collected sound samples. As shown in Fig. 5, the analysis of a large number of recordings collected at several locations, revealed that mechanical sounds are most prevalent between 1 and 2 kHz and biological sounds are most prevalent between 2 and 8 kHz (Gage and Napoletano, 2004; Gage et al., 2001).

To compute the overall level of biophony present in an acoustic signal, we first compute the power spectral density (PSD) (Welch, 1967) of the signal. Then, a rectangular estimate of the PSD integral is computed for the anthropogenic and biophonic frequency ranges, and NDSI computed as: $\mathrm{NDSI}=(\beta-\alpha)/(\beta+\alpha)$, where β and α are the total estimated PSD for the largest 1 kHz biophony bin and the anthrophony bin respectively. The NDSI for the soundscape at a location does not remain constant, and changes according to time of day or day of year can be used to plot how NDSI changes over time. Note that NDSI is a ratio in the range $[-1 \text{ to } +1]$, where $+1$ indicates a signal containing no anthrophony. However, for some biophonic vocalizations a low NDSI can also indicate the presence of certain types of animals. For instance, the common loon (*Gavia immer*) has a low frequency call that often produces a NDSI score less than -0.8. Such anomalies are expected when attempting to characterize the highly variable nature of soundscape acoustics, and indicates that advancements are needed to help characterize and search acoustic observations. However, searching by NDSI can be a useful filter to help limit the number of recordings for further examination.

4.2.2. Using the search bar

As shown in Fig. 2, a search bar is provided above the search results. Using the search bar, a library visitor can further refine the set of results returned after sensor unit, date, and time selection. By selecting a field (e.g., NDSI, Biophony, or a frequency bin) and a mathematical comparison operator, and then entering a value in the search field, acoustic observations can be retrieved that meet the specified frequency constraints. In addition, a Boolean search can be entered by typing a search string directly into the search field. For instance, to retrieve the subset of observations where the NDSI is less than -0.9 and the total power in the 1 kHz bin from 1 to 2 kHz is greater than 0.2, enter: $\$ndsi<-0.9\,and\,\$L1>0.2$. Power magnitudes (e.g., L1, L2 or Biophony) are vector normalized so that values are in the range $[0-1]$ to ease selection of search criteria. When entering a Boolean search, the drop-down selection boxes for field and operator selection are ignored.

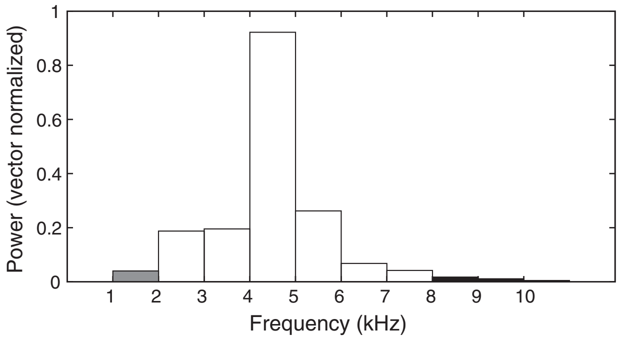

Fig. 5. Example 1 kHz frequency bins. Each bin is an estimate of the total power spectral density (PSD) for a specific 1 kHz range. Anthrophony is shown using gray fill and biophony using white fill. Bin values are vector normalized for plotting.

5. Symbolic representations

In this section, we first review methods used to represent and process acoustic data, including piecewise aggregate approximation (PAA) (Keogh et al., 2000; Yi and Faloutsos, 2000) and symbolic aggregate approximation (SAX) (Lin et al., 2003). Then we describe how a symbolic representation can be used to search the digital acoustic archive.

5.1. Piecewise aggregate approximation

Piecewise aggregate approximation (PAA) was introduced by (Keogh et al., 2000), and independently by (Yi and Faloutsos, 2000), as a means to reduce the dimensionality of time series. For completeness a brief overview of PAA is presented here; full details can be found in Keogh et al. (2000) and Yi and Faloutsos (2000). As shown in Figs. 6 and 7, an original time series sequence, Q, of length n is converted to PAA representation, $\overline{Q}$. First, Q is Z-normalized (Li and Porter, 1988) as follows: $\forall i \quad q_i = (q_i - \mu)/\sigma$, where μ is the vector mean of the original signal, σ is the corresponding standard deviation and q_i is the i^{th} element of Q. Second, Q is segmented into $w \leq n$ equal sized subsequences, and the mean of each subsequence computed. $\overline{Q}$ comprises the mean values for all subsequences of Q. Thus, Q is reduced to a sequence $\overline{Q}$ with length w. Each i^{th} horizontal segment of the plot shown in Fig. 7 represents a single element, q_i, of $\overline{Q}$. Thus, the complete PAA algorithm first Z-normalizes Q and then computes the segment means to construct $\overline{Q}$, as depicted in Fig. 7.

Z-normalization and conversion to PAA representation affords two benefits that facilitate unsupervised learning. First, clustering acoustic observations collected in natural environments is often impeded by variance in signal strength due to distance from the sensor station or differences between individual vocalizations. Z-normalization converts two signals that vary only in magnitude to two identical signals, enabling comparison of signals of different strength. Second, conversion to PAA representation helps smooth the original signal to facilitate comparison of recordings.

5.2. Symbolic aggregate approximation

Extending the benefits of PAA is a representation introduced by Lin et al. (2003) called Symbolic Aggregate approXimation (SAX). The purpose of SAX is to enable accurate comparison of time series using a symbolic representation. As shown in Fig. 8, SAX converts a sequence from PAA representation to symbolic representation, where each symbol (letter) appears with equal probability based on the assumption that the distribution of time series subsequences is Gaussian (Lin et al., 2003). Thus, each PAA segment is assigned a symbol by dividing the Gaussian probability distribution into α equally probable regions, where α is the alphabet size ($\alpha = 5$ in Fig. 8). Each PAA segment falls within a specific Gaussian region and is assigned the corresponding symbol.

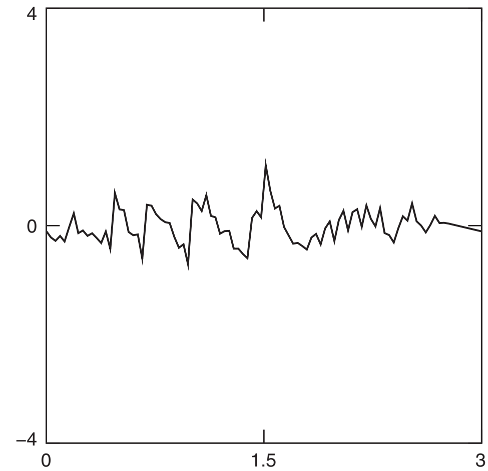

Fig. 6. Example original signal prior to Z-normalization and subsequent PAA processing.

5.3. Symbolic frequency search

Similar to the process for constructing a time series SAX representation, a 1 kHz symbolic signature was constructed for each acoustic observation, as shown in Fig. 9. First, a symbolic signature is constructed by Z-normalizing a vector comprising the 1 kHz bin PSD values for that recording. Each vector value is then assigned an alphabetic character based on the assumption that the distribution of PSD values (for a large number of recordings) is Gaussian. A visitor can retrieve similar results by clicking on the K SAX string (shown in Fig. 2). A search request can also be entered directly into the search bar. For instance, to search the Twin Lakes Soundscape data set for the call of the common loon (*G. immer*), a visitor could first select the entire Twin Lakes data set, and then enter the following search string: *$1ksax like jdddd%*, indicating that the subset of observations with a 1 kHz SAX string beginning with *jdddd* should be retrieved. Note, the entire data set is selected if all sensors, dates and times have been selected in the right hand panel (see Fig. 2). Specific bin values can also be omitted from the search by replacing a letter in the SAX string with an underscore.

6. Unsupervised learning

In this section, we first describe the hierarchical data clustering system we use for our studies on unsupervised learning using acoustics. As

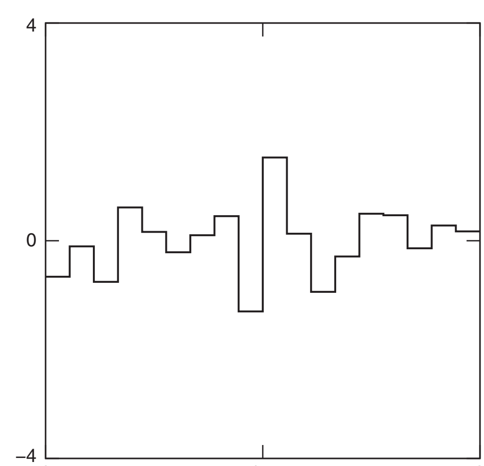

Fig. 7. Example signal after Z-normalization and subsequent PAA processing.

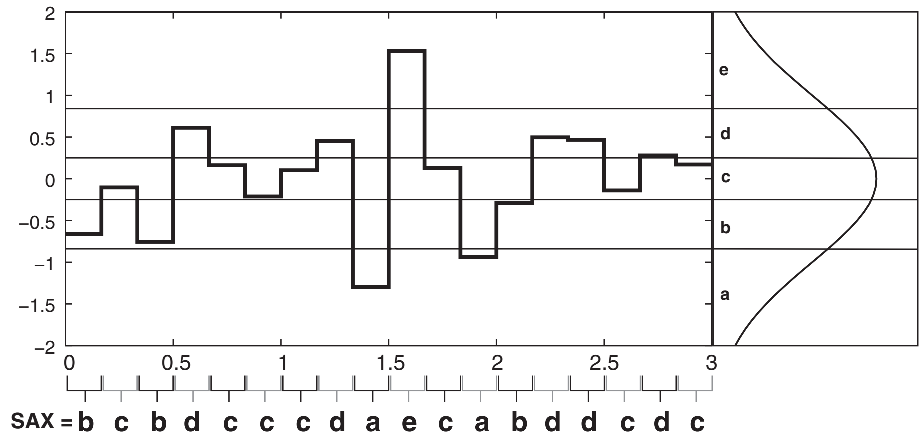

Fig. 8. Conversion of the example PAA processed signal to SAX.
Adapted from Lin et al. (2003).

mentioned earlier, unsupervised learning refers to techniques that group observations according to some metric of similarity without a priori categorical knowledge and is often called data clustering. Second, we present relevant background on the construction of SAX bitmaps that we later adapt to support visualization, clustering of acoustic observations, and searching of the digital library. Finally, we describe our adaptation of SAX bitmaps to support clustering and retrieval of library observations based on frequency space similarity.

6.1. MESO

For clustering acoustic observations we use MESO[1] (Kasten and McKinley, 2007; Kasten et al., 2010), a perceptual memory system designed to support online, incremental learning and decision making in autonomic computing systems. MESO is based on the well-known leader–follower algorithm (Hartigan, 1975), an online, incremental technique for clustering a data set. A novel feature of MESO is its use of small agglomerative clusters, called sensitivity spheres, that aggregate similar training patterns. Sensitivity spheres are partitioned into sets during the construction of a memory-efficient hierarchical data structure. This structure enables the implementation of a content-addressable perceptual memory system: instead of indexing by an integer value, the memory system is presented with a pattern similar to the one to retrieve from storage. MESO can be used strictly as a pattern classifier (Duda et al., 2001) if a categorization is known during training. In this case, each pattern is labeled, assigning each pattern to a specific real-world category. When evaluated on standard data sets, MESO accuracy compares very favorably with other classifiers, while requiring less training and testing time in most cases (Kasten and McKinley, 2007). However, MESO is not dependent on labeled data and can be used to cluster unlabeled data and construct a hierarchical model of the training data.

As shown in Fig. 10, two basic functions comprise the operation of MESO during data clustering: training and cluster path extraction. During training, patterns are stored in perceptual memory, enabling the construction of an internal model of the training data. Each unsupervised training sample is an unlabeled pattern x_i, where x_i is a vector of continuous, binary or nominal values. The size of the sensitivity spheres is determined by a δ value that specifies the sphere radius in terms of distance (e.g. Euclidean distance) from the sphere's center. Sensitivity sphere size is calculated incrementally, growing the δ during training.

Fig. 11 shows an example of sensitivity spheres for a 2D data set comprising three clusters. A sphere's center is calculated as the mean of all patterns that have been added to that sphere. The δ is a ceiling value for determining if a training pattern should be added to a sphere, or if creation of a new sphere is required.

Once MESO has been trained, a representation of the hierarchical structure produced can be extracted from MESO. For clustering acoustic observations, the representation extracted comprises the training pattern identifiers and associated cluster paths that indicate where each pattern is stored in the tree. For instance, as shown in Fig. 10, the cluster path 2030 is extracted, where each digit represents a step in the path down the tree that leads to the sphere that contains one or more training patterns. Each step in the path represents an increase in similarity (smaller distance) between a smaller set of training patterns. For instance, the cluster subpath 203 comprises a larger set of patterns that include all the patterns for path 2030 and other patterns that are somewhat less similar. Therefore, the desired level of similarity between patterns can be specified by selecting pattern subsets as represented by partial cluster paths. The branching factor of the MESO tree can also be specified. That is, if a branching factor of 2 is specified, each node will have up to 2 children, while a branching factor of 8 will produce up to 8 children per node. A branching factor of 8 will produce a wider, shallower tree than a branching factor of 2. In Section 8, we will plot results that show how the branching factor influences searching with cluster paths. For clustering with MESO, we use SAX bitmaps as a constant space representation of collected acoustic observations.

6.2. SAX bitmaps

Kumar et al. (2005) proposed time series bitmaps for visualization and anomaly detection in time series. SAX bitmaps are constructed by counting occurrences of symbolic subsequences of length n (e.g., 1, 2 or 3 symbols). Each bitmap can be represented using an n-dimensional matrix, where each cell represents a specific subsequence. An example is shown in Fig. 12; using subsequences of length $n=2$, matrix cell (a,a) contains the count and frequency with which the subsequence *aa* occurs. Frequencies are computed by dividing the subsequence count by the total number of subsequences.

We adapted SAX bitmaps to support frequency-based clustering and searching of library observations as follows. For each recording, the time varying spectral density is computed (spectrogram). Then, the set of spectral density values for each time step (spectrogram column) is converted to SAX representation. Finally, a SAX bitmap is constructed by counting the total number of occurrences of 3 letter subsequences for all columns and the rates of occurrence computed. In our study, we used an alphabet size of $n=4$ (a,b,c,d). The 3-dimensional matrix of subsequences is mapped to an 8×8 bitmap

[1] The term MESO refers to the tree algorithm used by the system (Multi-Element Self-Organizing tree).

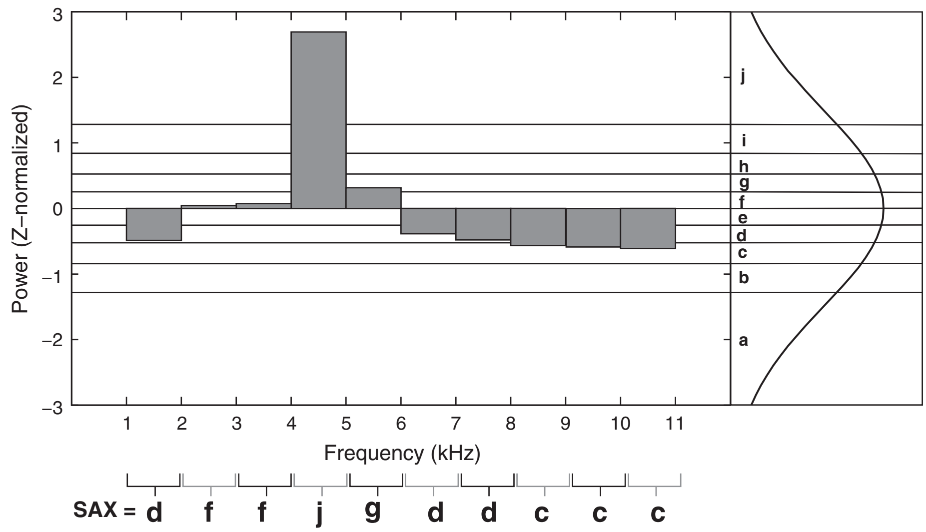

Fig. 9. Conversion of the *Z*-normalized example 1 kHz frequency bins shown in 5 to SAX format (SAX = dffjgddccc).

as shown in Table 1. As shown in Fig. 2, two bitmaps were constructed. The rates of occurrence were used directly to construct a color bitmap, where different colors indicate the rate of occurrence (lighter colors indicate higher rates). A black-and-white bitmap was constructed by coloring a bitmap cell white if the rate was greater than the per cell rate when subsequences are equally distributed ($\mu = \frac{1}{64}$) and colored black otherwise.

For clustering two vectors were constructed. The first vector was constructed using the rates of occurrence directly. The second vector comprises the binary values from the black-and-white bitmap. For all recordings in each library data set, a vector is constructed for each color and black-and-white bitmaps. These vectors are then used to train MESO. After training, the cluster paths are extracted and added to the database to support bitmap-based similarity searches. By clicking on either the color or black-and-white bitmap, a visitor can execute a search for other observations that have similar bitmaps. The search string used also appears in the search bar to enable editing of the cluster path and allow the visitor to retrieve a larger set of observations. For instance, after clicking on a color bitmap, the following search string might appear in the search bar: $cphex64 like 077321%. By removing the last digit from this string a shorter cluster path can be specified that extends to a shallower depth in the MESO tree. This new cluster path encompasses a larger number of sensitivity spheres that comprise a larger number of observations. By executing this new search, the visitor can retrieve a new, larger set of observations that are somewhat less similar than the original search.

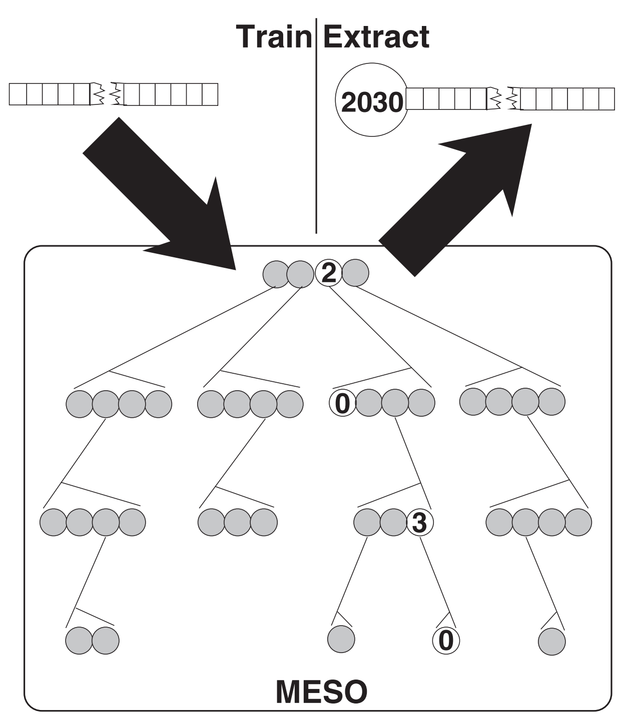

Fig. 10. The training of MESO using unlabeled patterns and the extraction of the cluster path associated with a specific training pattern. Each digit of the cluster path, 2030, represents the partition position of the sphere that contains the training pattern at each level of the tree.

7. Evaluation methods

In this section, we describe our approach for evaluating SAX strings and bitmaps for retrieval of similar bioacoustic recordings. The use of SAX strings and bitmaps for summarization enables users to use

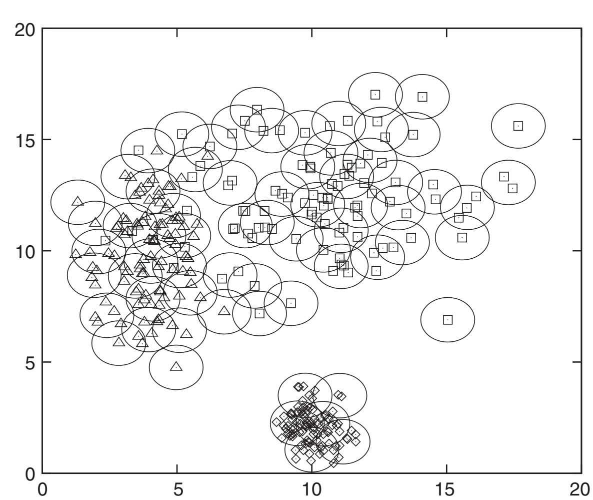

Fig. 11. MESO sensitivity spheres for three 2D-Gaussian clusters. Circles represent the boundaries of the spheres as determined by the current δ. Each sphere contains one or more training patterns, and each training pattern belongs to one of the three Gaussian clusters (diamond, square, or triangle).

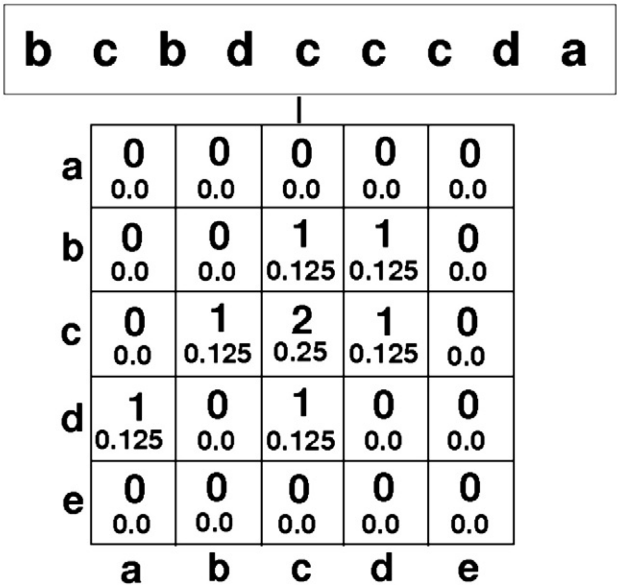

Fig. 12. Computing a SAX bitmap for a signal (see Kumar et al., 2005 for more information). Number of subsequences occurrences shown over frequency.

query-by-example for efficient search of the acoustic library. These summarized representations can be easily inserted into a relational database to support timely online retrieval. However, it is important to determine how well a summarized representation can be expected to perform as a search key. To evaluate the use of SAX strings and bitmaps for information retrieval, we will use clustering evaluation techniques to compare their performance against that of a baseline created using species census data created through expert interpretation of 228 one minute recordings. In general, data clustering is a method for organizing a data set into one or more groups (clusters) based on a feature vector (pattern) comprised of values that describe each element of the data set in a meaningful way. Typically, the relative similarity of one element to another is computed using a distance metric, such as Euclidean or Hamming distance. A more detailed discussion of the data sets, baseline, and a brief introduction to the evaluation metrics is presented below. For a more in depth discussion on clustering evaluation metrics, including mutual information and entropy, the reader is encouraged to see Section 3 and Manning et al. (2008).

7.1. Evaluation data set

The evaluation data set comprises 228 one minute recordings. Each recording was labeled according to the dominate vocalizing species heard within the recording. An integer value in the range 1 to 3 was used to indicate the abundance of species (small sized population, moderate size population, and full chorus, respectively). Although, labeling only indicates the dominate species and an estimate of abundance, other sounds are also found in these recordings (e.g., wind and other less dominate vocalizations). Typically, recordings collected by sensors in natural environments are noisy and highly variable. Table 2 is a list of the species labels, including the number of occurrences within the data set and a species code used in later discussion.

In addition to the SAX strings and bitmaps, a census pattern was constructed for each recording. A census pattern is a vector of integer values that represent the abundance of each species. The field of the dominate vocalizing species is set to 1, 2, or 3, depending on its abundance, all other fields are set to 0 (species is not dominate). This data set is used as a baseline for comparison with our SAX representations to better understand how these automated summarization techniques compare with manual interpretation. Notably, our goal is not species recognition but an understanding of how well these synoptic representations capture the information required to relate similar recordings. Shown in Fig. 13, are two polar dendrograms constructed using average link clustering and Euclidean distance with a randomly selected 100 pattern subset of the full data set. A visual comparison reveals that census patterns cluster into a small number of clusters comprised of like species. The SAX patterns also tend to cluster like species together, but into a larger number of clusters with a dendrogram of greater depth. Although this visualization is informative, it is not sufficient to fully clarify how well the SAX representations will support query-by-example retrieval. Next, we will introduce two distance metrics, two information-theoretic, and two decision-based metrics that we will use to further evaluate the potential of the SAX representations.

7.2. Distance measures

To evaluate the use of SAX strings and bitmaps for clustering acoustic observations, we use two distance metrics: Hamming distance and Euclidean distance. Hamming or edit distance is the count of differences between two patterns. That is, if two patterns differ at location i, the distance between these two patterns is incremented by 1, otherwise the distance remains unchanged. Intuitively, Euclidean distance is the length of the line segment between two patterns. Formally, Euclidean distance is defined as:

$$distance_{Euclidean}(Q,P) \equiv \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}, \quad (1)$$

where Q and P are two patterns of length n.

7.3. Evaluation metrics

7.3.1. Purity

Purity is a measure of the homogeneity of the data clusters produced by a clustering algorithm. To compute purity, each cluster is labeled as belonging to the most common pattern class assigned to that cluster. For example, the first cluster in Fig. 14 would be assigned to class BCCH and the second to class RASY. The number of patterns in each class that

Table 1
Positions of sequences of length 3 in a bitmap with an alphabet size of 4.

aaa	aab	aba	abb	baa	bab	bba	bbb
aac	aad	abc	abd	bac	bad	bbc	bbd
aca	acb	ada	adb	bca	bcb	bda	bdb
acc	acd	adc	add	bcc	bcd	bdc	bdd
caa	cab	cba	cbb	daa	dab	dba	dbb
cac	cad	cbc	cbd	dac	dad	dbc	dbd
cca	ccb	cda	cdb	dca	dcb	dda	ddb
ccc	ccd	cdc	cdd	dcc	dcd	ddc	ddd

Table 2
Major vocalizing species identified in the acoustic data set used for evaluating the clustering potential of the summarizing metrics.

Code	Occur	Common name (*scientific name*)
AMCR	8	American crow (*Corvus brachyrhynchos*)
BCCH	15	Black-capped chickadee (*Poecile atricapillus*)
BLJA	7	Blue jay (*Cyanocitta cristata*)
CANG	41	Canada goose (*Branta canadensis*)
CHSP	20	Chipping sparrow (*Spizella passerina*)
COLO	26	Common loon (*Gavia immer*)
PSCR	35	Northern spring peeper (*Pseudacris crucifer*)
RACL	19	Green frog (*Rana clamitans*)
RASY	25	Wood frog (*Rana sylvatica*)
RWBL	2	Red-winged blackbird (*Agelaius phoeniceus*)
TIXX	30	Dog day cicada (*Tibicen* spp.)

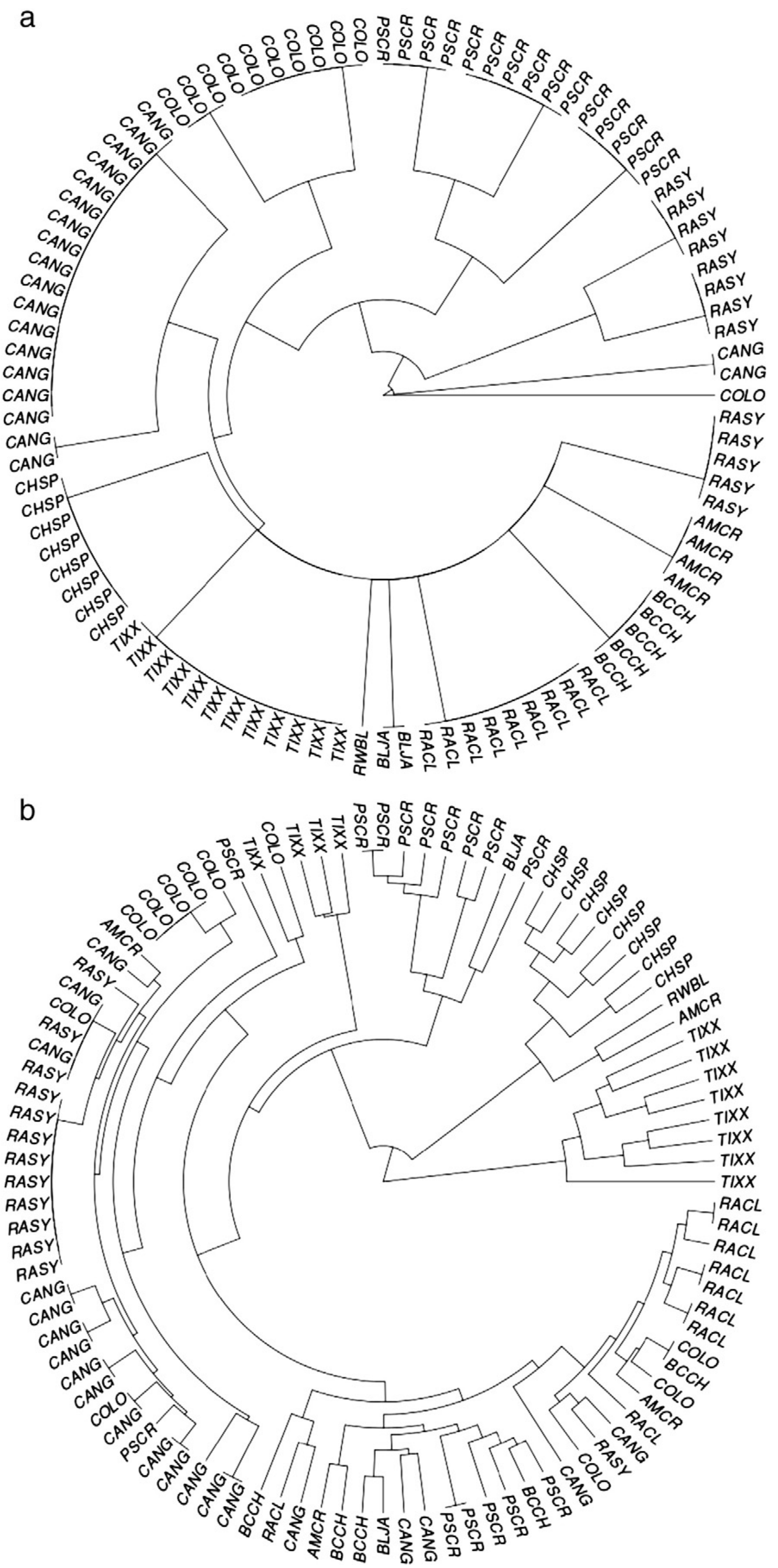

Fig. 13. Polar dendrograms for a randomly selected subset of 100 acoustic observations: a) a dendrogram constructed using Euclidean distance and census pattern features (baseline), b) a dendrogram using Euclidean distance and 1 kHz SAX pattern features with an alphabet of size 10.

belong to that cluster's class label are summed and divided by the total number of patterns. Formally, purity is defined as:

$$purity(\Omega, \mathbb{C}) = \frac{1}{N}\sum_{k} \max_{j} \left|w_k \cap c_j\right|, \tag{2}$$

where $\Omega = \{w_1, w_2, \ldots, w_k\}$ is the set of clusters, $\mathbb{C} = \left\{c_1, c_2, \ldots, c_j\right\}$ is the set of species classes, and N is the total number of patterns clustered.

Purity is a measure that ranges from [0,1], where 0 indicates a low level of purity, where patterns are well mixed between classes. A purity of 1 indicates that patterns have been assigned only to clusters with the same species class. Referring to Fig. 14, we can compute the purity of these two clusters by observing that the largest intersection for the left and right clusters occurs for BCCH (3) and RASY (3), respectively. Therefore, the purity of this clustering is $6/10 = 0.60$ ($N = 10$).

7.3.2. Normalized mutual information

Intuitively, normalized mutual information (NMI) measures the increase in information afforded by a particular set of clusters and normalizes this measure to the range [0,1]. Formally, NMI is defined as:

$$NMI(\Omega, \mathbb{C}) = \frac{I(\Omega, \mathbb{C})}{[H(\Omega) + H(\mathbb{C})]/2}, \tag{3}$$

where I is mutual information and H is entropy. Mutual information is defined as:

$$I(\Omega, \mathbb{C}) = \sum_{k}\sum_{j} \frac{\left|w_k \cap c_j\right|}{N} \log_2 \frac{N\left|w_k \cap c_j\right|}{\left|w_k\right|\left|c_j\right|}, \tag{4}$$

and cluster entropy is defined as:

$$H(\Omega) = -\sum_{k} \frac{|w_k|}{N} \log_2 \frac{|w_k|}{N}, \tag{5}$$

and similarly class entropy is defined as:

$$H(\mathbb{C}) = -\sum_{j} \frac{\left|c_j\right|}{N} \log_2 \frac{\left|c_j\right|}{N}. \tag{6}$$

An NMI of 0 indicates that no information has been gained, and that patterns are essentially randomly assigned to different clusters, while an NMI of 1 indicates that the patterns have been assigned to clusters that exactly recreate the species classes. The NMI of the clustering shown in Fig. 14 can be computed by observing that the cardinality of each cluster is 5, and that the cardinality of each species class is 3, 1, 4, and 2 for BCCH, CHSP, RASY and RACL, respectively. The cardinality of the intersection between each cluster and each species class can also be computed for BCCH, CHSP, RASY and RACL as 3, 1, 1 and 0 (left cluster) and 0, 0, 3, 2 (right cluster). These cardinalities can then be inserted into Eqs. (4), (5), (6) and (3) to compute $I(\Omega, \mathbb{C}) = 0.68$, $H(\Omega) = 1$, $H(\mathbb{C}) = 1.85$ and $NMI(\Omega, \mathbb{C}) = 0.47$.

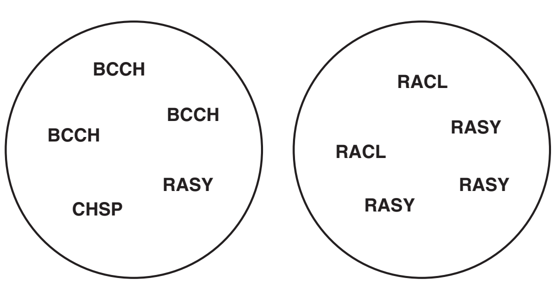

Fig. 14. Example clusters showing two clusters comprised of four species of birds and amphibians. Examples described in the discussion refer to this figure. More in depth examples can be found in Manning et al. (2008).

7.3.3. Rand index

Intuitively, the Rand index (RI) is a measure of accuracy that measures the number of correct classifications as a rate that ranges between [0,1]. Computing RI requires counting the number of pattern pairs that have the same label for each cluster. That is, if two patterns belonging to the same cluster have the same label, then they are considered to be correctly classified. If only one pattern with a particular label is present in a cluster, it is considered as incorrectly classified. For instance, using Fig. 14, the true positive count (TP) can be computed directly by counting the number of like labeled pattern pairs for each cluster or by using combinations:

$$TP = \binom{3}{2} + \binom{3}{2} + \binom{2}{2} = 7, \tag{7}$$

where TP is the number of patterns that are *true positives* ($TP = 7$ for Fig. 14). The first combination in Eq. (7) corresponds with BCCH in the left cluster and the second and third combinations correspond with

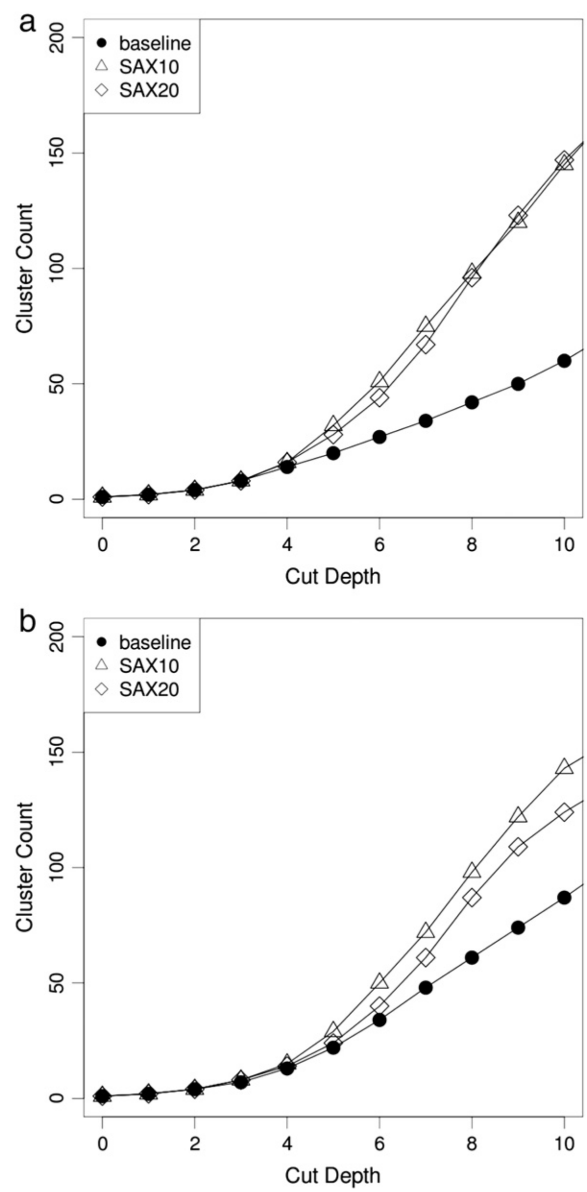

Fig. 15. Number of clusters produced using Hamming distance and Euclidean distance for the baseline and 1 kHz SAX strings with alphabets of sizes 10 and 20.

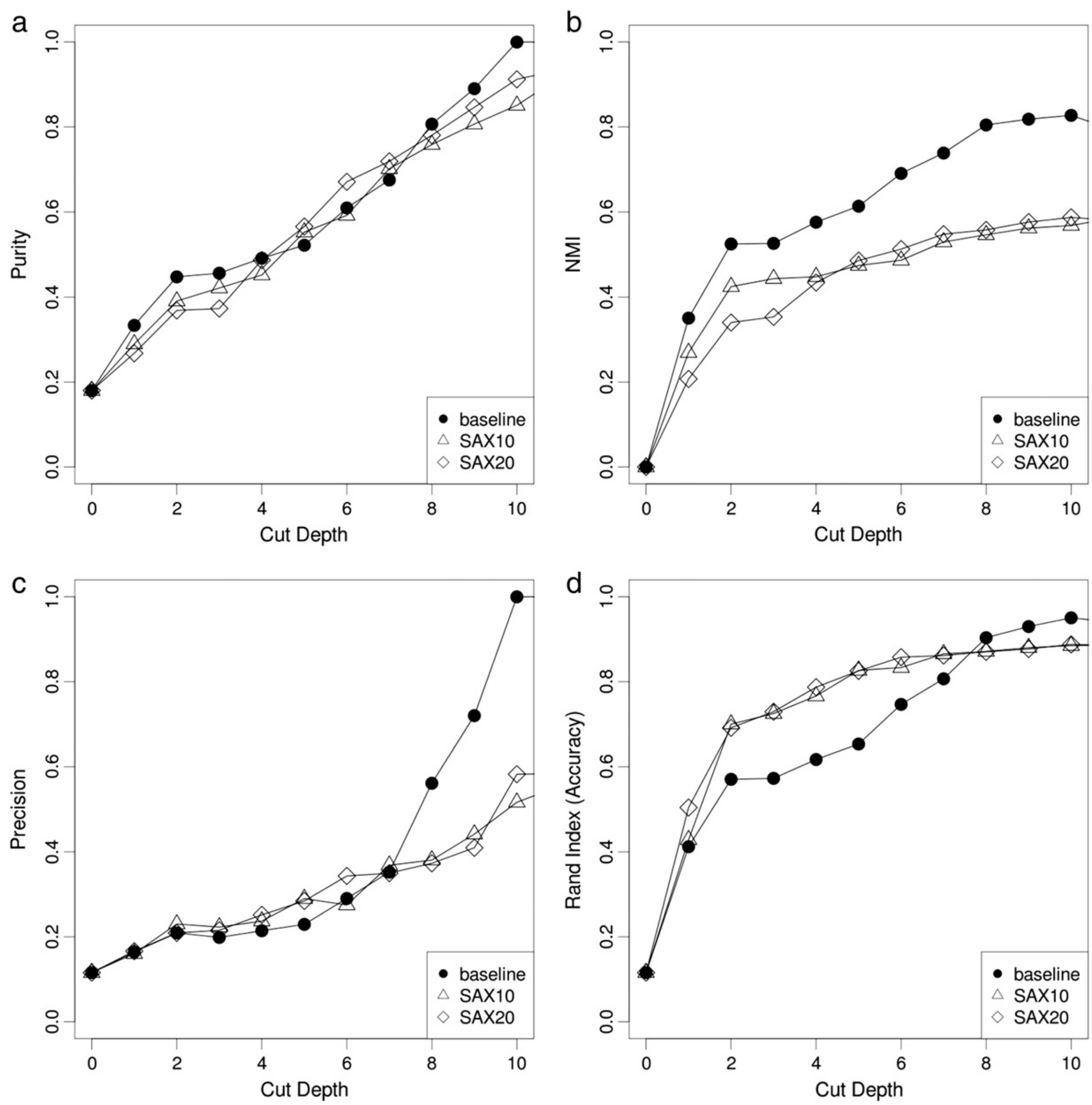

Fig. 16. Purity, NMI, precision and RI cluster metrics when using Hamming distance with 1 kHz SAX pattern features. Baseline included for comparison.

RASY and RACL in the right cluster. A true positive is a pattern with a specific species class that has been classified as such. Similarly, a *false positive* is a pattern that is classified as a species that differs from its true species class. *True negatives* and *false negatives* are defined similarly to true positives and false positives. We can calculate the true negatives (*TN*), false positives (*FP*), and false negatives (*FN*) as follows. First, compute the total number of true and false positives, by summing the total number of pairs for each cluster:

$$TP + FP = \binom{5}{2} + \binom{5}{2} = 20. \tag{8}$$

Subtracting the true positives (Eq. (7)) from the result of Eq. (8) gives us the false positives (13 pairs). The total number of pairs for all 10 patterns can be then computed as follows:

$$TP + FP + TN + FN = \binom{10}{2} = 45, \tag{9}$$

and subtracting the result of Eq. (8) from (9) gives us $TN + FN$ (25 pairs). To compute the false negatives, we need to compute the total number of pairs for each class (BCCH,CHSP,RASY,RACL) and subtract the class-wise true positives as follows:

$$FN = \left[\binom{3}{2} - \binom{3}{2}\right] + \left[\binom{4}{2} - \binom{3}{2}\right] + \left[\binom{2}{2} - \binom{2}{2}\right] = 5, \tag{10}$$

where the first, second and third bracketed terms represent the pairwise computation for classes BCCH, RASY and RACL, respectively (CHSP has no pairs and is omitted). To compute the true negatives we subtract the result of Eq. (10) from $TN + FN$ (20 pairs). Using the results of preceding equations, RI can be formally defined as:

$$RI = \frac{TP + TN}{TP} + FP + FN + TN = 0.6. \tag{11}$$

Although computing the combinations required by RI could be prohibitive for large data sets, the following closed form reduces the computational load significantly:

$$\binom{n}{2} = \frac{n(n-1)}{2}. \tag{12}$$

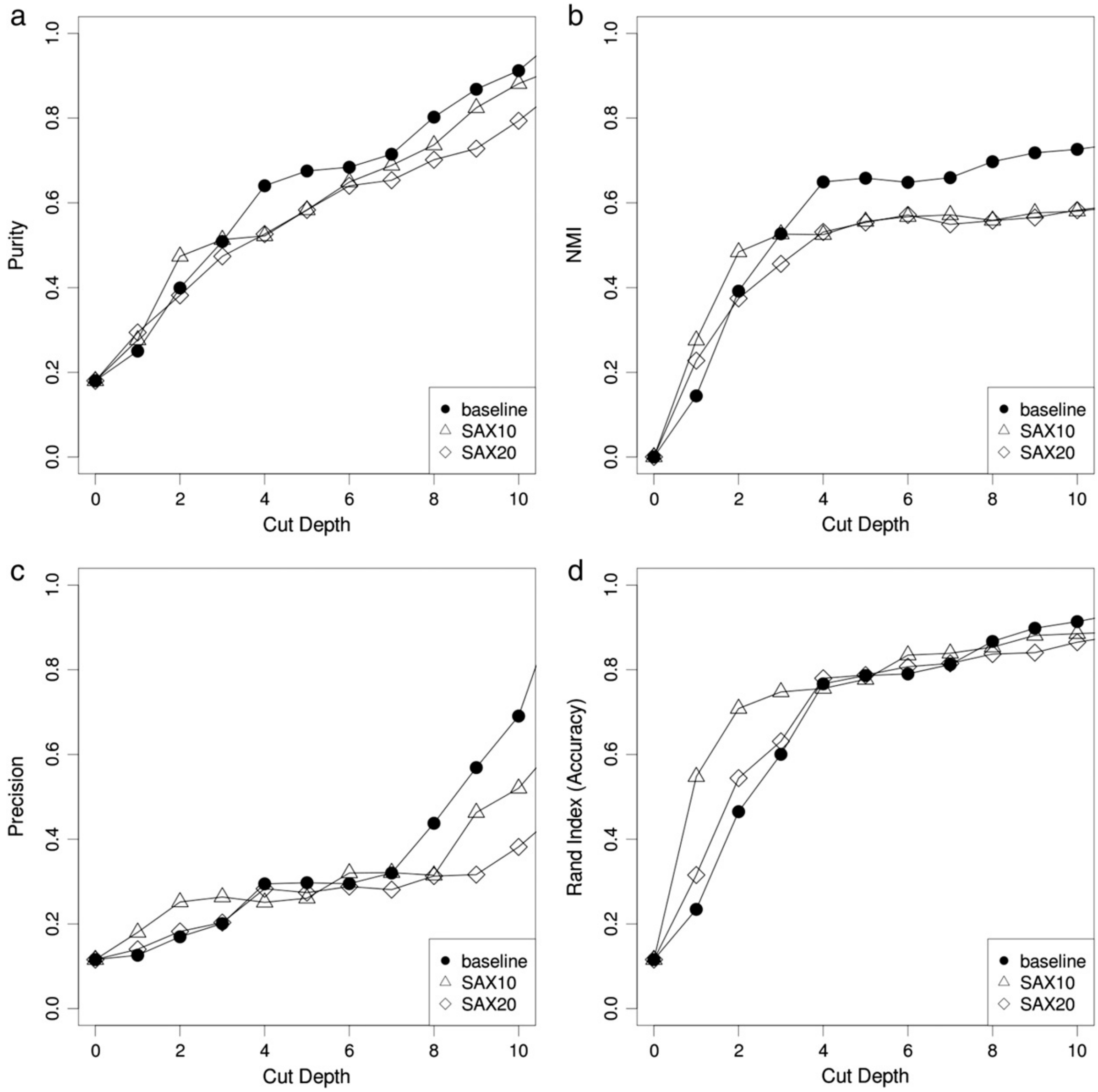

Fig. 17. Purity, NMI, precision and RI cluster metrics when using Euclidean distance with 1 kHz SAX pattern features. Baseline included for comparison.

7.3.4. Precision

Intuitively, precision is the probability that a randomly chosen pattern is correctly clustered together, or classified, with other patterns with the same species label. Precision is also known as the true positive rate. Precision is defined as:

$$P = \frac{TP}{TP} + FP, \qquad (13)$$

where TP and FP are the number of true and false positives, respectively.

8. Experimental results

In this section, we present results using the cluster metrics described in Section 7. We plot these results against dendrogram cut depth to show how the metrics change as clusters become smaller and more numerous at greater depths. For SAX strings, we consider SAX alphabets of size 10 and 20 to better understand the effect of alphabet size on cluster quality. For SAX bitmaps, we plot how the metrics change with MESO branching factor, helping to understand if smaller branching factors are more favorable than the larger ones. We provide baseline plots using census patterns for comparing SAX summary representations with manual, expert interpretation. First, we will consider how SAX strings compare with the baseline, and then we will investigate SAX bitmaps. It is worth noting that although we have labeled the recordings to indicate the predominate vocalizing species, our goal is not specific to species identification but to better understand how our synoptic representations enable retrieval and organization of similar recordings. The baseline is provided for comparison with the results of a manual survey.

As shown in Fig. 15, the number of clusters produced using the baseline census patterns is lower than that for SAX strings with an alphabet size of 10 or 20. This is consistent with what is shown using the polar dendrograms in Fig. 13. This is as expected since representations, such as SAX strings and bitmaps, strive to summarize

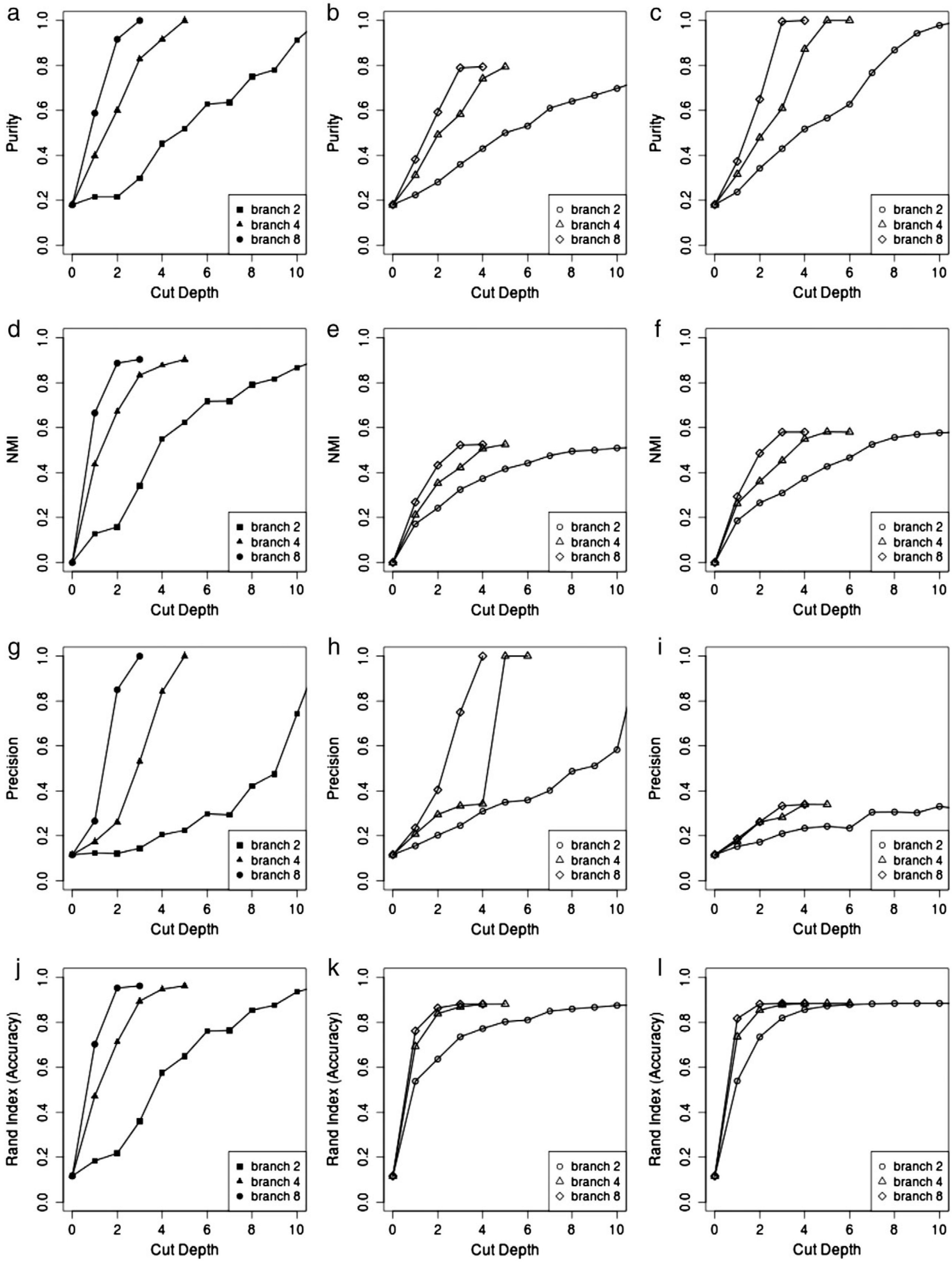

Fig. 18. Purity, NMI, precision and RI cluster metrics when using Euclidean distance with species census patterns (column 1), black-and-white SAX bitmaps (column 2), and color SAX bitmaps (column 3).

more complex representations, but do not reduce the recordings to basic census information. Notably, the baseline with Euclidean distance produces more clusters than that for Hamming distance.

8.1. Evaluation of SAX strings

Fig. 16 plots purity, NMI, precision and RI cluster metrics against the baseline for SAX strings with alphabet sizes of 10 and 20 using Hamming distance. For each metric, both SAX strings have a similar response, while the baseline for NMI and precision shows a notable improvement over the SAX representations at greater cut depths. Notably, this improvement is likely due, in part, to the baseline dendrogram having a shallower depth than that produced for the SAX string representations. The similarity of response between the baseline and the SAX string representations indicates that SAX strings are capturing important characteristics of the recordings despite reduction to a much smaller characterization than that afforded by the raw recordings.

Fig. 17 plots purity, NMI, precision and RI cluster metrics against the baseline for SAX strings with alphabet sizes of 10 and 20 using Euclidean distance. In general, these plots show a greater similarity between the SAX string representations and the baseline. This is likely due to the Euclidean distance producing more clusters at increasing cut depths than the Hamming distance. However, it is notable that the use of Euclidean distance is encouraging a convergence between the baseline and the SAX string representations. As such, the SAX string representations show promise as useful summarizations of acoustic recordings for organizing and searching large data sets.

8.2. Evaluation of SAX bitmaps

Fig. 18 depicts the plots for purity, NMI, precision and RI for the species census and abundance patterns (column 1) for comparison with the plots of these metrics for black-and-white and color SAX bitmaps shown in columns 2 and 3, respectively. In general, all metrics improve more rapidly as the MESO branching factor increases, indicating, as expected, that a larger branching factor will produce more rapid improvement of cluster quality along shorter cluster paths. As shown in columns 2 and 3 of Fig. 18, a larger branching factor has a similar effect when using black-and-white or color bitmaps.

The cluster metrics for using black-and-white bitmaps are plotted in the second column of Fig. 18. The plots for black-and-white bitmaps are similar to those for species census patterns (Fig. 18 column one), but attain lower values, indicating that there is some loss of information when summarizing acoustic recordings using black-and-white bitmaps. Notably, for a branching factor of 8, census patterns attain approximate values of: 1.0, 0.9, 1.0 and 0.96 for purity, NMI, precision and RI, respectively. These values compare with those for black-and-white bitmaps of: 0.8, 0.53, 0.34 and 0.88, indicating a significant degradation in NMI and precision, while purity and RI only reduced by 0.2 and 0.08, respectively.

Shown in the third column of Fig. 18 are the cluster metrics for using color bitmaps. Again, the plots are similar to those in column one of Fig. 18, but attain somewhat lower values. However, color bitmaps also show some improvement over black-and-white bitmaps, indicating that they capture more of the information found in the acoustic recordings than black-and-white bitmaps. For a branching factor of 8, the purity, NMI, precision and RI values for color bitmaps are: 1.0, 0.58, 1.0 and 0.88, respectively. Notably, there is only slight improvement in NMI over black-and-white bitmaps (0.05), and their RIs are equal. Purity and precision are equal to those for species census patterns.

8.3. Correlative summary

Table 3 shows the Pearson's correlation between each summarization method and the species census patterns for each of the four metrics, and the number of clusters produced when using SAX bitmaps. Pearson's correlation ranges from -1 to $+1$, with $+1$ indicating a strong positive correlation and -1 indicating a strong negative correlation. In most cases, a strong positive correlation exists for each metric. However, only a weak positive correlation exists for precision when using SAX strings due to commission errors when using this method. In general, the plots for the summarized acoustic representations correlate well with those for the census patterns, but would likely return more unwanted results during execution of a search than would be returned using census patterns. However, this is expected since summarization of acoustic recordings elides some information in order to ease organizing and searching large archives. Moreover, summarization can proceed automatically without manual, human interpretation that would be time prohibitive for large numbers of recordings.

Table 3
Pearson's correlation for cluster metrics between the baseline and summarization methods.

	Clusters	NMI	Precision	Purity	RI
Hamming distance					
SAX10	0.92	0.80	0.44	0.99	0.95
SAX20	0.91	0.80	0.09	0.99	0.94
Euclidean distance					
SAX10	0.96	0.91	0.38	0.99	0.93
SAX20	0.97	0.93	0.11	0.98	0.99
Cluster paths (black and white)					
Branch 2	na	0.97	0.85	0.99	0.85
Branch 4	na	0.99	0.94	0.99	0.94
Branch 8	na	0.97	0.98	0.97	0.99
Cluster paths (color)					
Branch 2	na	0.98	0.99	0.98	0.76
Branch 4	na	0.99	0.93	0.98	0.92
Branch 8	na	0.97	0.92	0.95	0.98

9. Conclusions and future work

We have described the current status of the REAL digital library and our use of temporal, geographic, symbolic and cluster-based search techniques. Our evaluation of these techniques indicates that representations that summarize acoustic recordings can enable improved organization and searching of large acoustic archives. As such, these archives will better support scientific inquiry, by allowing users to distill the data down to relevant subsets for addressing specific questions. In addition, a small set of manually identified recordings of interest can be used as keys for query-by-example searches of large archives that return larger sets of pertinent acoustic observations.

Based on our investigations, we intend to build on this proof-of-concept system and extend our studies by adding more depth and rigor to our investigation of symbolic and machine learning techniques to better understand and evaluate their application for searching and organizing large sensor data repositories. In our study, we used cluster evaluation metrics that relied on each pattern having a single, discrete label. However, environmental sound recordings may contain the vocalizations of more than one species. It is our intent to investigate evaluation metrics that can encompass datasets of observations with multiple labels.

We have discussed techniques for organizing and searching acoustic recordings collected in natural settings. Sensed acoustic data is a highly complex and dense data type that presents significant processing challenges due to both size and complexity. The extraction of interesting signals and the characterization of recordings in a meaningful way can enable wider use of acoustics for monitoring and managing the ecosystems around us. It is our intent to develop algorithms and methods that ease interpretation of acoustic and other complex sensor data types, and provide public access to methods and data through our web site. It is our hope that our approach to open access to sensor data will help promote research and education at all levels.

Acknowledgments

This research was supported in part by a Research Excellence Fund grant from Michigan State University and by the Kellogg Biological Station, a National Science Foundation long term ecological research site. The authors would like to thank Evan Bowling at Michigan State University for their contributions to this work. We also thank the reviewers for their helpful comments and suggestions.

References

Achour, M., Betz, F., Dovgal, A., Lopes, N., Magnusson, H., Richter, G., Seguy, D., Vrana, J., 2009. PHP Manual.

Adams, M., Cox, T., Moore, G., Croxford, B., Refaee, M., Sharples, S., December 2006. Sustainable soundscapes: noise policy and the urban experience. Urban Studies 43 (13), 2385–2398.

Anderson, R., Morrison, M., Sinclair, K., Davis, H., Kendall, W., December 1999. Studying wind energy/bird interactions: a guidance document. Prepared for the National Wind Coordinating Committee.

Barber, J., Crooks, K., Fristrup, K., 2009. The costs of chronic noise exposure for terrestrial organisms. Trends in Ecology & Evolution 25, 180–189.

Berndt, D.J., Clifford, J., July 1994. Using dynamic time warping to find patterns in time series. Proceedings of KDD-94: AAAI Workshop on Knowledge Discovery in Databases, Seattle, Washington, USA, pp. 359–370.

Bystrak, D., 1981. The North American breeding bird survey. Studies in Avian Biology 6, 34–41.

Canadian Wildlife Service, July 2006. Recommended Protocols for Monitoring Impacts of Wind Turbines on Birds.

Carles, J.L., Barrioa, I.L., de Lucio, J.V., 1999. Sound influence on landscape values. Landscape and Urban Planning 43, 191–200.

Chou, C.-H., Lee, C.-H., Ni, H.-W., September 2007. Bird species recognition by comparing the HMMs of the syllables. 2nd International Conference on Innovative Computing, Information and Control. Kumamoto, Japan.

Cornell Laboratory of Ornithology, 2009. Macaulay library. (last date accessed: 20 September 2009). http://www.macaulaylibrary.org.

Department of Evolution, Ecology and Organismal Biology, 2009. Borror laboratory of bioacoustics. (last date accessed: 10 October 2009). http://blb.biosci.ohio-state.edu/.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification, Second edition. John Wiley and Sons, Incorporated, New York, New York, USA.

Estrin, D., Michener, W., Bonito, G., August 2003. Environmental cyberinfrastructure needs for distributed sensor networks: a report from a national science foundation sponsored workshop. Tech. rep., Scripps Institute of Oceanography, 12 May 2005. www.lternet.edu/sensor_report/.

Fagerlund, S., 2007. Bird species recognition using support vector machines. EURASIP Journal on Advances in Signal Processing (2007, 8 pages).

Fagerlund, S., Härmä, A., September 2005. Parameterization of inharmonic bird sounds for automatic recognition. 13th European Signal Processing Conference (EUSIPCO). Antalya, Turkey.

Flanagan, D., 2006. JavaScript: The Definitive Guide. O'Reilly Media, Inc., Sebastopol, California.

Florida Museum of Natural History, 2009. Ornithology. (last date accessed: 10 October 2009). http://www.flmnh.ufl.edu/natsci/ornithology/ornithology.htm.

Gage, S.H., Napoletano, B.M., 2004. Envirosonics Equipment Manual. Computational Ecology and Visualization Laboratory. Michigan State University, East Lansing, Michigan, USA.

Gage, S.H., Napoletano, B., Cooper, M.C., 2001. Assessment of ecosystem biodiversity by acoustic diversity indices. Journal of the Acoustical Society of America 109 (5), 2430.

Gannon, J.E., 1974. Twin Lakes status. Tech. Rep. 1. University of Michigan Biological Station.

Hannah, L., Lohse, D., Hutchinson, C., Carr, J., Lankerani, A., 1994. A preliminary inventory of human disturbance of world ecosystems. Ambio 23, 246–250.

Hartigan, J.A., 1975. Clustering Algorithms. John Wiley and Sons, Inc., New York, New York, USA.

Hobson, K., Rempel, R., Greenwood, H., Turnbull, B., Wilgenburg, S., 2002. Acoustic surveys of birds using electronic recordings: new potential from an omnidirectional microphone system. Wildlife Society Bulletin 30, 709–720.

Hutto, R.L., Pletschet, S.M., Hendricks, P., 1986. A fixed-radius point count method for nonbreeding and breeding season use. Auk 103, 593–602.

Joo, W., 2008. Environmental sounds as an ecological variable to understand dynamics of ecosystems. Master's thesis, Department of Zoology, Michigan State University, East Lansing, Michigan.

Joo, W., Gage, S.H., Kasten, E.P., 2011. Analysis and interpretation of variability in soundscapes along an urban–rural gradient. Landscape and Urban Planning 103, 259–276. http://dx.doi.org/10.1016/j.landurbplan.2011.08.001.

Kasten, E.P., McKinley, P.K., April 2007. MESO: supporting online decision making in autonomic computing systems. IEEE Transactions on Knowledge and Data Engineering (TKDE) 19 (4), 485–499.

Kasten, E.P., McKinley, P.K., Gage, S.H., June 2007. Automated ensemble extraction and analysis of acoustic data streams. Proceedings of the 1st International Workshop on Distributed Event Processing, Systems and Applications (DEPSA), held in conjunction with the 27th IEEE International Conference on Distributed Computing Systems (ICDCS). Toronto, Ontario, Canada.

Kasten, E.P., McKinley, P.K., Gage, S.H., 2010. Ensemble extraction for classification and detection of bird species. Ecological Informatics 5, 153–166.

Ke, Y., Hoiem, D., Sukthankar, R., June 2005. Computer vision for music identification. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. San Diego, California, US.

Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S., 2000. Dimensionality reduction for fast similarity search in large time series databases. Knowledge and Information Systems 3 (3), 263–286.

Kogan, J.A., Margoliash, D., 1998. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: a comparative study. Journal of the Acoustical Society of America 103 (4), 2185–2196.

Krause, B., Gage, S.H., Joo, W., 2011. Measuring and interpreting the temporal variability in the soundscape at four places in Sequoia National Park. Landscape Ecology 26 (9), 1247–1256.

Krausman, P., Leopold, B., Scarbrough, D., 1986. Desert mule deer response to aircraft. Wildlife Society Bulletin 14, 68–70.

Kumar, N., Lolla, N., Keogh, E., Lonardi, S., Ratanamahatana, C.A., April 2005. Time-series bitmaps: a practical visualization tool for working with large time series databases. Proceedings of SIAM International Conference on Data Mining (SDM'05). Newport Beach, California, USA, pp. 531–535.

Li, K.-P., Porter, J.E., April 1988. Normalizations and selection of speech segments for speaker recognition scoring: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)., 1, pp. 595–598.

Lin, J., Keogh, E., Lonardi, S., Chiu, B., June 2003. A symbolic representation of time series with implications for streaming algorithms. Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. San Diego, California, USA.

Manning, C.D., Raghavan, P., Schütze, H., 2008. An Introduction to Information Retrieval. Cambridge University Press, Cambridge, England.

Mellinger, D.K., Clark, C.W., June 2000. Recognizing transient low-frequency whale sounds by spectrogram correlation. Journal of the Acoustical Society of America 107 (6), 3518–3529.

Mellinger, D.K., Clark, C.W., 2006. MobySound: a reference archive for studying automatic recognition of marine mammal sounds. Applied Acoustics 67 (11–12), 1226–1242.

Michener, W.K., Baerwald, T., Firth, P., Palmer, M.A., Rosenberger, J., Sandlin, E., Zimmerman, H., 2001. Defining and unraveling biocomplexity. BioScience 51, 1018–1023.

Michigan Department of Labor and Economic Growth, December 2005. Michigan Siting Guidelines for Wind Energy Systems.

Neumann, P., Merriam, H., 1972. Ecological effects of snowmobiles. Canadian Field-Naturalist 86, 209–212.

O'Connor, R.J., Dunn, E., Johnson, D.H., Jones, S.L., Petit, D., Smith, C.R., Welling, E., February 2000. A Programmatic Review of the North American Breeding Bird Survey: Report of a Peer Review Panel.

Pijanowski, B.C., Farina, A., Gage, S.H., Dumyahn, S.L., Krause, B.L., 2011a. What is soundscape ecology? An introduction and overview of an emerging new science. Landscape Ecology 1–20.

Pijanowski, B.C., Villanueva-Rivera, L.J., Dumyahn, S.L., Farina, A., Krause, B.L., Napoletano, B.M., Gage, S.H., Pieretti, N., March 2011b. Soundscape Ecology: The Science of Sound in the Landscape.

Porter, J., Arzberger, P., Braun, H.-W., Bryant, P., Gage, S., Hansen, T., Hanson, P., Lin, C.-C., Lin, F.-P., Kratz, T., Michener, W., Shapiro, S., Williams, T., July 2005. Wireless sensor networks for ecology. Bioscience 55 (7), 561–572.

Qi, J., Gage, S.H., Joo, W., Napoletano, B., Biswas, S., 2008. Soundscape characteristics of an environment: a new ecological indicator of ecosystem health. In: Ji, W. (Ed.), Wetland and Water Resource Modeling and Assessment. CRC Press, New York, New York, USA, pp. 201–211. Ch. 17.

Ralph, C.J., Sauer, J.R., Droege, S., 1995. Monitoring Bird Populations by Point Counts. United States Department of Agriculture.

Romano, T., Keogh, M., Kelly, C., Feng, P., Berk, L., Schlundt, C., Carder, D., Finneran, J., 2004. Anthropogenic sound and marine animal health: measures of the nervous and immune systems before and after intense sound exposure. Canadian Journal of Fisheries Aquatic Science 61, 1124–1134.

Rosenstock, S.S., Anderson, D.R., Giesen, K.M., Leukering, T., Carter, M.F., 2002. Landbird counting techniques: current practices and an alternative. Auk 119, 46–53.

Schafer, R.M., 1977. The Tuning of the World. Alfred Knopf, New York, New York, USA.

Schafer, R., 1994. The Soundscape: Our Sonic Environment and the Tuning of the World. Destiny Books, Rochester, Vermont.

Skole, D.L., Batzli, S., Gage, S.H., Pijanowski, B., Chomentowski, W., Rustem, W.R., 2002. Forecast Michigan: Tracking Change for Land Use Policy and Decision Making. Institute for Public Policy and Social Research, Michigan State University.

Somervuo, P., Härmä, A., May 2004. Bird song recognition based on syllable pair histograms. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Montreal, Quebec, Canada.

Southworth, M., 1969. Sonic environments of cities. Environment and Behavior 1 (1), 49–70.

Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S., December 2008. Rapid acoustic survey for biodiversity appraisal. PLoS One 3 (12), e4065.

Sun Microsystems, 2009. MySQL 5.1 Reference Manual. Sun Microsystems. Santa Clara, California.

The 2020 Science Group, June/July 2005. Towards 2020 Science. Report from the Towards 2020 Science Workshop.

The Apache Software Foundation, 2009. Apache HTTP Server Reference Manual.

The British Library, 2009. Archival sound recordings: environment and nature. (last date accessed: 10 October 2009). http://sounds.bl.uk/.

Thompson, W.L., 2002. Towards reliable bird surveys: accounting for individuals present but not detected. Auk 119, 18–25.
Thompson, J.N., Reichman, O.J., Morin, P.J., Polis, G.A., Power, M.E., Sterner, R.W., Couch, C.A., Gough, L., Holt, R., Hooper, D.U., Keesing, F., Lovell, C.R., Milne, B.T., Molles, M.C., Roberts, D.W., Strauss, S.Y., 2001. Frontiers of ecology. BioScience 51, 15–24.
Truax, B. (Ed.), 1978. The World Soundscape Project's Handbook for Acoustic Ecology. A.R.C. Publications, Vancouver, B.C.
Truax, B. (Ed.), 1984. Acoustic Communication. Ablex Publishing, Norwood, New Jersey.
Truax, B. (Ed.), 1999. Handbook for Acoustic Ecology, second edition. Cambridge Street Publishing, Vancouver, B.C.
United States Fish and Wildlife Service, May 2003. Service interim guidance on avoiding and minimizing wildlife impacts from wind turbines. Memorandum to regional directors.
Vilches, E., Escobar, I.A., Vallejo, E.E., Taylor, C.E., August 2006. Data mining applied to acoustic bird species recognition. Proceedings of the International Conference on Pattern Recognition (IPCR). Hong Kong, China, pp. 400–403.
Vitousek, P., Mooney, H., Lubchenco, J., Melillo, J., 1997. Human domination of earth's ecosystems. Science 277, 494–499.
Warren, P., Katti, M., Ermann, M., Brazel, A., 2006. Urban bioacoustics: it's not just noise. Animal Behaviour 71, 491–502.
Weir, L., Mossman, M., 2005. North American Amphibian Monitoring Program (NAAMP). University of California Press, Berkeley, California, USA.
Welch, P.D., June 1967. The use of the fast Fourier transform for the estimation of power spectra: a method based on time-averaging over short, modified periodograms. IEEE Transactions on Audio and Electroacoustics AU-15, pp. 70–73.
West, B.W., Flikkema, P., Sisk, T., Koch, G.W., August 2001. Wireless sensor networks for dense spatio-temporal monitoring of the environment: a case for integrated circuit, system and network design. IEEE Circuits and Systems Workshop on Wireless Communications and Networking. University of Notre Dame, Indiana, USA.
Wildlife Acoustics, 2009. http://www.wildlifeacoustics.com (last date accessed: 25 April 2011).
Wrightson, K., 2000. An introduction to acoustic ecology. Soundscape 1, 10–13.
Yi, B.-K., Faloutsos, C., September 2000. Fast time sequence indexing for arbitrary Lp norms. Proceedings of the 26th International Conference on Very Large Databases. Cairo, Egypt.