RE-SONIFICATION OF GEOGRAPHIC SOUND ACTIVITY USING ACOUSTIC, SEMANTIC, AND SOCIAL INFORMATION

Alex Fink, Brandon Mechtley, Gordon Wichern, Jinru Liu, Harvey Thornburg, Andreas Spanias, and Grisha Coleman

School of Arts, Media and Engineering, SenSIP Center,
& School of Electrical, Computer and Energy Engineering
Arizona State University
alex.fink@asu.edu, bmechtley@asu.edu

ABSTRACT

Sonic representations of spaces have emerged as a means to capture and present the activity that conventional representations, such as maps, do not encapsulate. Therefore, to convey the activity information of regions, both large and small, we use sounds and information provided by regional communities in the automated design of soundscapes to re-sonify geographic sound activity. To quantify this community knowledge, we have developed an ontological framework to determine the importance of sound and concepts to one another using acoustic, semantic, and social information. This framework is then used in the automated design of a generative soundscape model purposed to identify and re-sonify sounds that impart relevant information about a geographic region. Furthermore, we are developing a social networking website to facilitate the collection and re-sonification of sounds and data.

1. INTRODUCTION

The ability to understand the activity local to specific geographic regions is limited when presented through maps, directories, and other conventional representations. Even novel interactive representations, such as Photosynth [1] and Google Street View [2], present community information in the form of artifacts (in this instance, images). Exploring geography through sonification, however, presents a method of experiencing a location through sonic events. This concept has been explored in a number of systems that primarily focus on displaying where sounds are recorded and allowing them to be played as recorded [3, 4, 5]. As a primary carrier of information about activity, sound can project information about how people relate to their surrounding environments and what these environments mean in terms of their daily lives. This link between geography and activity is often explored in the context of soundscapes. As acoustic experiences are considered to play a significant role in human ecology, soundscapes may supplement our comprehension of activity in physical environments and geographic spaces as well as our understanding of cultural and anthropological issues [6, 7, 8]. Whether real or imagined, soundscapes have been used to enhance immersive experiences in real and virtual worlds for purposes including music [9], audio-visual production [10], geographic exploration [11, 12, 13], and community understanding [14]. Many previous innovative works in soundscape synthesis address community meaning and aesthetics in interactive systems through the knowledge of a composer, often gained through community presence, interaction, and/or interviews [11, 12, 13, 15]. We seek, however, to create scalable, automated methods of soundscape design, where meaning is defined by communities themselves, drawing on their provision of both sound recordings and community knowledge.

Meaningful re-sonification of activity in a geographic region can be difficult when recorded sounds from that region are either 1) abundant or 2) scarce. Where recordings in a region are few in number, re-sonification itself may be sparse or highly repetitive without the inclusion of relevant sounds from other locations. Conversely, if recordings from a region are plentiful, many sounds may be redundant or uninformative about the area's activity. Both situations may be addressed by classifying and using those sounds that are relevant and important to an area. Traditional classification of sounds within a soundscape (keynote, signal, and soundmark) is primarily focused on their perceptual role to listeners [6, 7]. This classification is area-specific, depending on the perception of sounds as dictated by the meaning and prevalence of sounds in a community. While the identification of important sounds to an area does not provide this classification, it is able to distinguish which sounds convey the relevant activity of a region, a relevance perhaps best determined by that region's own community.

The concept of community-defined importance of sounds has long been held in the auditory field; in [6], Schafer states,

> Acoustic design should never become design control from above. It is rather a matter of the retrieval of a *significant aural culture*, and that is a task for everyone.

This idea also extends beyond the auditory domain; Google's PageRank technology, for example, determines the importance of web pages by considering the number and relative importance of other pages that link to them [16]. The relevance of such pages is then defined by the internet community's own activity. Similarly, the acoustic knowledge and the actions of a community can help to reveal important sounds for the re-sonification of geographic activity.

To work towards revealing this importance, we have developed an ontological framework to link sounds together through acoustic, semantic, and social information. Using acoustic content in conjunction with user-provided tags, our framework relies on the prior knowledge of acoustic and semantic ontologies combined with community-defined social links between sounds

This material is based upon work supported by the National Science Foundation under Grant No. 0504647.

and concepts. By linking concepts and sounds together with the community-provided information, the ontological framework provides a measure of the relevance of sounds (and concepts) to one another. To re-sonify specified locations through the playback of sounds in a database, the ontological framework is used to create a graph-based generative soundscape model. Similar to the use of textual queries to filter a ranked list of important websites, we use location to determine the soundscape model parameters such that geographically relevant sounds play frequently. Consideration of the size (surface area) of locations allows our re-sonification to scale to communities or regions of varying size. Using sounds recorded from these locations and other locations that are deemed important by an area's community, our methodology aims to create meaningful soundscapes reflective of the geographic sound activity in those areas. User-tests to assess our methodology are needed, though informal reviews by users have thus far been generally favorable.

The remainder of the paper is organized as follows. Section 2 describes our ontological framework to link sounds and concepts together, using acoustic, semantic, and social information. The application of this framework to the automated design of a soundscape model for re-sonifying geographic activity is discussed in Section 3. Section 4 then presents a social networking website currently under development that provides for the collection of and classification of sounds; the site features an interactive map using our re-sonification scheme to allow virtual "soundwalks." Finally, preliminary results are given in Section 5, followed by conclusions and discussion of future work in Section 6.

2. ONTOLOGICAL FRAMEWORK

To automatically compose soundscapes from collections of sounds with user-provided descriptions, some notion of similarity between sounds is necessary to determine what sounds may be relevant to a space. For example, if few sounds are recorded in a location, retrieving perceptually similar sounds provides greater diversity in the synthesis process. We calculate such similarity with an ontological framework that links together sounds and concepts, using acoustic similarity between sounds, social information in the form of links between sounds and concepts, and semantic information in the form of conceptual similarity [17]. Using these separate modalities, the ontological framework determines the relevancy of objects (sounds and concepts) to one another, using available links (acoustic, social, or semantic).

The ontological framework consists of an undirected graph (Figure 1), where nodes in the graph represent sounds ($\mathcal{S} = \{s_1, ...s_N\}$) or concepts ($\mathcal{C} = \{c_1, ...c_M\}$). Nodes are connected by weighted links, and a nonnegative link weight connecting nodes i and j is signified by $W(i, j)$. Links of weight zero represent equivalence between the nodes connected by that link, while a link of infinite weight between two nodes is equivalent to no link being present. Given a subset of nodes, $\mathcal{A}$, and query node, q, a posterior distribution from the network can be calculated as follows:

$$P(a \in \mathcal{A}|q) = \frac{e^{-d^*(q,a)}}{\sum_{b\in\mathcal{A}} e^{-d^*(q,b)}}, \quad (1)$$

where $d^*(q, a)$ is the shortest-path distance in the network between nodes q and a, which can be efficiently computed using Dijkstra's algorithm [18]. Note that, in the case of a query node that does not yet exist in the database, such as a new sound or concept, the distances between the query node and all other nodes of its type can be computed on demand. For example, when a new sound is uploaded to the database ($q \in \mathcal{S}$) and the ontological framework returns a distribution over concepts ($\mathcal{A} \subset \mathcal{C}$) as in Figure 1 (a), we can automatically annotate a new sound file with tags suggested by the community based on the audio content. In a similar fashion, a concept query ($q \in \mathcal{C}$) can return a distribution over sounds ($\mathcal{A} \subset \mathcal{S}$) as in Figure 1 (b). In order to use the ontological framework in this fashion, we must set the values for all link weights. A description of the three types of links we use, sound-to-sound, concept-to-concept, and sound-to-concept, follows below.

2.1. Acoustic information: sound-to-sound links

Sound-to-sound weights can be computed by comparing the acoustic content of each sound. This process begins with acoustic feature extraction, where six low-level features are calculated using a frame-based analysis, where we use 40 ms frames with 50% overlap and a Hamming window. Features are calculated either from the time-domain data or the short-time Fourier Transform (STFT) spectrum. The feature trajectory for a sound file is given by $Y_{1:T}^{(1:P)}$ where $Y_t^{(i)}$ is the ith feature value at frame t.

The six features we use include *loudness*, the dB-scaled RMS level over time; temporal sparsity, the ratio of ℓ^∞ and ℓ^1 norms calculated over all short-term RMS levels computed in a one-second interval; spectral sparsity, the ratio of ℓ^∞ and ℓ^1 norms calculated over the STFT magnitude spectrum; bark-weighted *spectral centroid*, a measure of the mean frequency content for a sound frame; transient index, the ℓ^2 norm of the difference of Mel frequency cepstral coefficients (MFCCs) between consecutive frames; and harmonicity, a probabilistic measure of whether or not the STFT spectrum for a given frame exhibits a harmonic frequency structure. For more details on how these features are calculated, see [19]. This feature set was developed to accurately represent a broad range of environmental sounds rather than any specific class of sounds (e.g. speech or music) while also providing an intuitive and minimal set for efficient retrieval of sounds stored in a database. To compare sounds, [20] describes a method of estimating $L(s_i, s_j) = \log P[Y_{1:T}^{(1:P)}(s_i)|\lambda^{(1:P)}(s_j)]$, the log-likelihood that the feature trajectory of sound s_i was generated by the hidden Markov Model (HMM) $\lambda^{(1:P)}(s_j)$ built to approximate the simple feature trends of sound s_j.

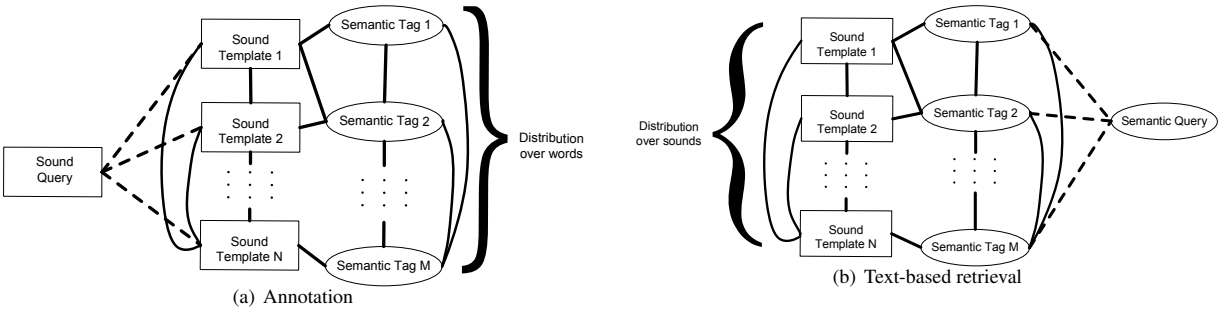
The ontological framework we have defined is an undirected, acyclic graph, which requires weights be *symmetric* ($W(s_i, s_j) = W(s_j, s_i)$) and *nonnegative* ($W(s_i, s_j) \geq 0$). Therefore, we cannot use the log-likelihood $L(s_i, s_j)$ as the link weight between nodes s_i and s_j, because it is not guaranteed to be symmetric and nonnegative. Fortunately, a well known semi-metric that satisfies these properties and approximates the distance between HMMs exists [17, 21]. Using this semi-metric we define the link weight between nodes s_i and s_j as

$$W(s_i, s_j) = \frac{1}{T_i}[L(s_i, s_i) - L(s_i, s_j)] + \frac{1}{T_j}[L(s_j, s_j) - L(s_j, s_i)], \quad (2)$$

where T_i and T_j represent the length of the feature trajectories for sounds s_i and s_j, respectively.

2.2. Semantic information: concept-to-concept links

To calculate concept-to-concept link weights, we use a similarity metric from the WordNet::Similarity library [22]. Specifically,



(a) Annotation (b) Text-based retrieval

Figure 1: Organization of ontological framework for two common indexing and retrieval tasks. Dashed lines indicate links added at query time. Text-based retrieval use words to search for (unlabeled) sounds. Annotation automatically describes a sound file based on its audio content and provides suggested tags to contributors.

we use the `vector` metric, because it supports the comparison of adjectives and adverbs, which are commonly used to describe sounds. The `vector` metric computes the co-occurrence of two concepts within the collections of words used to describe other concepts (their *glosses*) [22]. For a full review of WordNet similarity, see [23, 22].

By defining $Sim(c_i, c_j)$ as the WordNet similarity between the concepts represented by nodes c_i and c_j, an appropriately scaled link weight between these nodes is

$$W(c_i, c_j) = -\log\left[\frac{Sim(c_i, c_j)}{\max_{k,l} Sim(c_k, c_l)}\right]. \quad (3)$$

2.3. Social information: sound-to-concept links

We quantify the social information connecting sounds and concepts using a $M \times N$ dimensional votes matrix V, with elements V_{ji} equal to the number of users who have tagged sound s_i with concept c_j divided by the total number of users who have tagged sound s_i. By appropriately normalizing the votes matrix, it can be interpreted probabilistically as

$$P(s_i, c_j) = V_{ji} / \sum_k \sum_l V_{kl} \quad (4)$$

$$P(s_i|c_j) = V_{ji} / \sum_k V_{jk} \quad (5)$$

$$P(c_j|s_i) = V_{ji} / \sum_k V_{ki}, \quad (6)$$

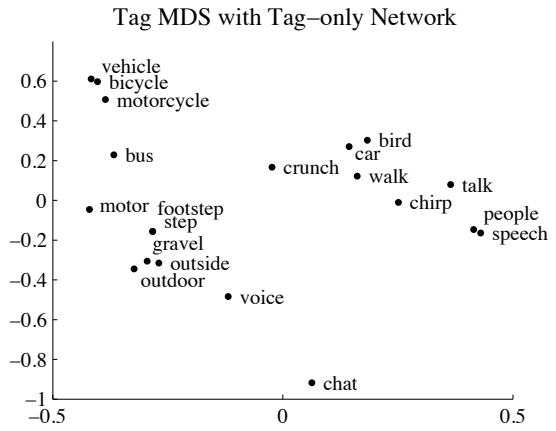
where $P(s_i, c_j)$ is the joint probability between s_i and c_j, $P(s_i|c_j)$ is the conditional probability of sound s_i given concept c_j, and $P(c_j|s_i)$ is defined similarly. Our goal in determining the social link weights connecting sounds and concepts is that the probability distributions output by the ontological framework using (1) are as close as possible to the conditional distributions from the votes matrix in (5) and (6). One way of measuring the distance between probability distributions is the Kullback-Leibler divergence [24]. The link weights between sounds and concepts are then optimized to jointly minimize the Kullback-Leibler divergence between the distributions obtained from the ontological framework and those from the votes matrix, using each sound in the database to obtain a distribution over concepts and each concept in the database to obtain a distribution over sounds. Complete details on this weight optimization process are provided in [17]. Empirically, we have found that a simple approximation of the optimized weight values is to set them to a value inversely related to the joint distribution (4), i.e., $W(s_i, c_j) = -\log P(s_i, c_j)$.

Presently, the votes matrix is obtained using only a simple tagging process. In the future we hope to augment the votes matrix with other types of community activity, such as discussions, rankings, or page navigation paths on a website. Furthermore, sound-to-concept link weights can be set as compositional parameters rather than learned from a "training set" of tags provided by users. For example, sounds can be made equivalent to certain emotional concepts (happy, angry, etc.) through the addition of zero-weight connections between specified sounds and concepts. These emotional connections will then affect the display and soundscape resynthesis processes discussed in subsequent sections. Similarly, relative scalings of weights between different types of information (e.g., semantic versus acoustic) can be used to explore different relationships amongst the collected sounds.

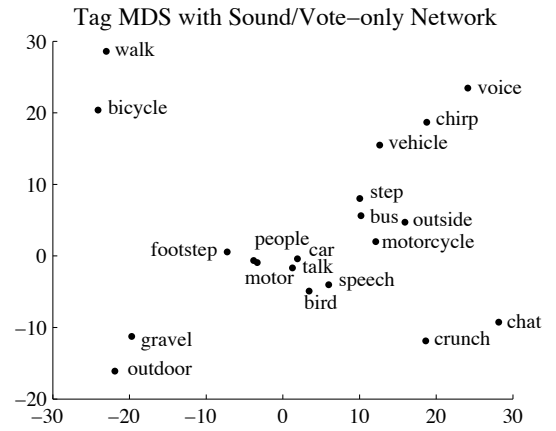
2.4. Multidimensional scaling

In order to conveniently summarize the social, semantic, and acoustic information contained in the ontological framework for soundscape re-synthesis and visual representation of sound activity on a map, we use multidimensional scaling (MDS) [25]; this embed each sound or concept node in the graph into a low-dimensional space in such a way that retains the distance relationships between nodes. MDS operates on a distance matrix, which is obtained by finding the shortest-path distance between all node pairs using Dijkstra's algorithm.

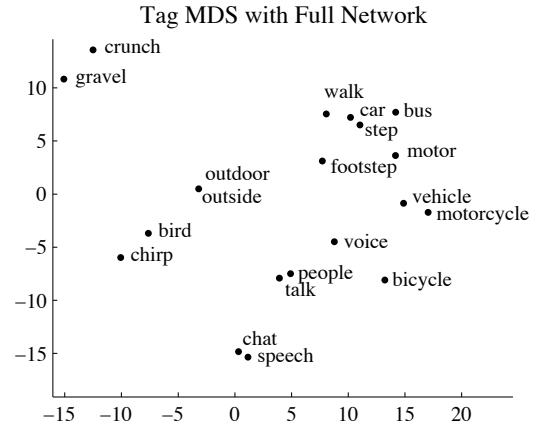
To provide an example of how the MDS embeddings of our ontological framework represent social, semantic, and acoustic information, Figures 2(a), 2(b), and 2(c) display the two-dimensional MDS for a subset of selected tags. In Figure 2(a) the distance matrix is calculated from an ontological framework containing only tag nodes, i.e., only semantic information, while Figure 2(b) contains both sound and tag nodes but only uses acoustic and social links, excluding concept-to-concept semantic connections. Figure 2(c) shows the MDS that uses all available nodes and links, i.e., acoustic, social, and semantic information. The differences between the absolute scales of the axes in the figures result from the different distance matrices, but by comparing relative tag positions we can see how information is organized in the different frameworks. From Figure 2(a), we can see that natural clusters



(a)Network containing only semantic weights



(b)Network containing only social and acoustic weights



(c)Network containing semantic, social, and acoustic weights

Figure 2: MDS of tags for networks with and without non-semantic information using the `vector` semantic similarity metric.

form from the semantic information, such as {*vehicle*, *bicycle*, *motorcycle*, *bus*}. Similarly, synonyms such as *outside*/*outdoor* are near each other. However, some concepts we might expect to group together do not, e.g., *chat*/*speech*. Similarly, in Figure 2(b), concepts such as *car* and *motor* are close, but *bird* and *chirp* are not.

By including social and acoustic information in the framework, as in Figure 2(c), the concepts organize into clusters that are informed by which concepts sound alike. For example, *chat*/*speech* are now quite near each other. Similar new clusterings can be seen between word pairs such as *gravel*/*crunch* and *bird*/*chirp*. Some clusterings are more vague in the reorganization, such as that which formerly grouped all vehicles, as we are now also capturing information of what concepts typically are heard together or in similar circumstances. This clustering behavior is then considered in our method of soundscape synthesis, described in the following section.

3. SOUNDSCAPE DESIGN

In the ontological framework, the relevance and importance of different sounds to one another is quantified. Using this information, we have developed a method to automatically design graph-based generative soundscape models to re-sonify geographic sound activity. This methodology aims to provide automated soundscape design that 1) is scalable to geographic regions of any size, and 2) meaningfully incorporates community knowledge to address re-sonification when locally obtained sounds are scarce or overly plentiful. Our design and synthesis of soundscapes assumes the availability of a database of sounds with GPS locations and community-provided tags; our process of collecting this data is presented in Section 4.1. Figure 3 displays how the components of our soundscape synthesis system interact, described as follows in detail.

3.1. Markov Transition Networks

Many of the recent approaches to soundscape generation use probabilistic generative models to sonify activity, focusing generally on designing the overall distribution of sounds in the soundscape. By using a stochastic generative model, all generated soundscapes provide different experiences that are consistent in their meaning, yet unique in each instance. In [10], for example, a method of ambiance generation is developed, where short isolated sounds are mixed with longer atmospheric recordings to generate a soundscape described by user-provided textual queries. The addition of the short, isolated sounds to the mix is determined probabilistically such that they most often appear in periods of relative silence in the overall mix. Other methods, such as that of [15], generate a soundscape from the randomized playback of pre-classified sounds where the expected temporal density of different types of sounds is determined through user interaction and design. Recent work [12, 13] uses a manually designed graph-based model where sound sequences are determined stochastically but subject to the set of sequences allowed by the graph's design. This model is particularly useful in modeling representations of complex sources of sound events. Drawing upon this and other work, we use graph-based modeling for soundscape synthesis, but with additional focus on the overall expected temporal density of sounds.

To generate soundscapes for our application, we have chosen to use an emerging compositional structure that we call a Markov Transition Network (MTN), a variation of the models introduced

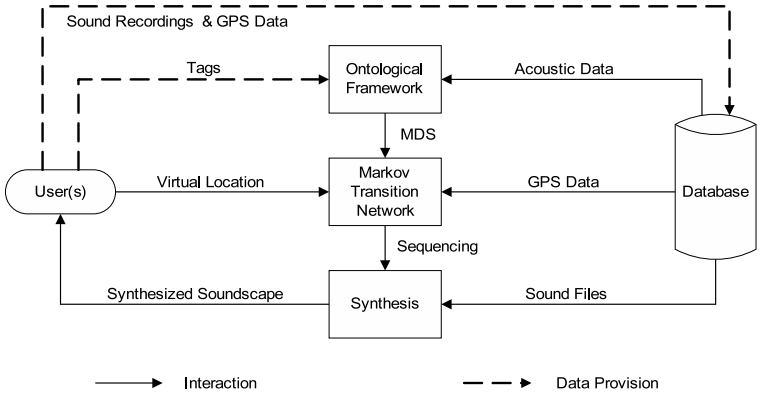


Figure 3: Diagram of the soundscape synthesis system.

in [12, 13, 26]. An MTN is a directed graph with N nodes, with possible directed edges from each node, i, to another node, j, including $j = i$. Figure 4 displays and example MTN. The MTN is used by an *actant process*, $A(t)$, that "travels" to the various nodes of the network, with its behavior dictated by the edges present in the graph. The actant process takes on values from 1 to N, representing the node at which the process is located at a given point in time. Each edge has an associated transition time, $\Delta(i, j)$. When $A(t)$ "enters" node i, the choice of the "next node," j, is determined by an associated probability, $P(i, j)$, and the actant process waits a time of $\Delta(i, j)$ before making the transition. If no edge exists between any two nodes, the associated probability is zero. Given these properties, we note that $A(t)$ is not a Markov process, as transition times depend on the origin and destination nodes, though it is a Semi-Markov process with deterministic transition times. Figure 4 displays an example MTN, with nodes and transition times of edges labeled. (Edge probabilities are omitted for clarity.)

Sound synthesis is performed by the sequenced playback of sounds as determined by the actant process. Sounds in the database are uniquely associated with a node, i, and a duration, $D(i)$. Upon $A(t)$ reaching a new node, the associated sound is played back in full, regardless of the chosen transition time to the next node or length of the following sounds. We presently mix together all sounds being played back into a single soundscape, though we note that a more complex multi-channel scheme could be adopted, and various effects (e.g., reverb) may be applied to individual sounds or the entire mix. Note that multiple actant processes may be active at any time, independently triggering sounds.

Using an MTN, the sequencing of sounds is made random, but it may be limited by the connections made between nodes. If only a single edge is directed from a node, then the sequencing upon the actant process's selection of that node will be temporarily deterministic. However, if all nodes in an MTN are fully connected, the behavior of the actant process becomes less predictable (dependent on the transition probability distributions). By limiting the number of edges connecting nodes, the sequencing determined by actant processes may be made variable, yet confined by the parameters of the network. This is considered in [12, 13], where limited connections are made between clusters of nodes to specify the behavior of complex sources of sound as predictable sequences. We recognize this effect of limiting connections, but we also wish

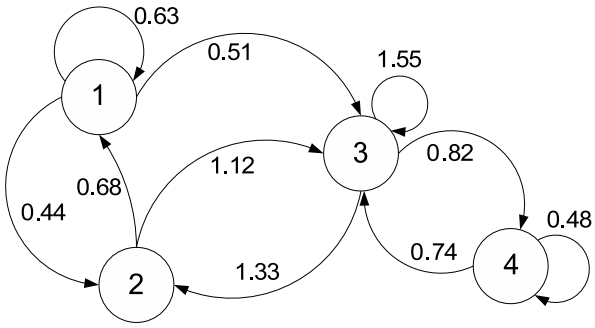


Figure 4: Example MTN for soundscape synthesis. Edges are labeled with transition times. Transition probabilities are not shown.

to examine the overall expected properties of the synthesized output. Therefore, we consider the expected temporal density of all available sounds.

For a sound i, with an intensity value (this may be any meaningful chosen measure, such as loudness), $V(i)$, we define the expected sum of intensities of any instances of sound i at a given time to be the density, $Density(i)$, given by

$$Density(i) \quad = \quad \frac{D(i)V(i)}{T(i,i)}, \tag{7}$$

where $T(i, j)$ is the expected time for the actant process to travel from node i to node j, including indirect paths. ($T(i, j)$ is then the expected time for the actant process to leave and return to node i). If the actant process travels directly to j, the transition time will simply be the delay, $\Delta(i, j)$, else it will be the delay, $\Delta(i, k)$, to an intermediary node, k, and the time taken to then reach j. Therefore,

$$T(i,j) \quad = \quad \sum_{k=1}^{N} P(i,k)\Delta(i,k) + \sum_{k=1,k\neq j}^{N} P(i,k)T(k,j). \tag{8}$$

Letting Δ, P, and T be $N \times N$ matrices with with elements $\Delta(i, j)$, $P(i, j)$, and $T(i, j)$, respectively, we may express (8) in matrix-vector form as

$$T(:,j) \quad = \quad C + Q_j T(:,j), \tag{9}$$

where $T(:,j)$ is the j^{th} column of T, the i^{th} element of the $N \times 1$ vector, C, is

$$C(i) \quad = \quad \sum_{k=1}^{N} P(i,k)\Delta(i,k), \tag{10}$$

and Q_j is P with the j^{th} column zeroed out. This gives

$$T(:,j) \quad = \quad (I - Q_j)^{-1}C, \tag{11}$$

which may be iterated over j.

While this allows us to analyze the density of sounds in a soundscape, for the purpose of design, we seek the ability to specify network parameters to create a desired density of sounds (this density may be determined by the interaction to which the resulting soundscape is applied). As the available sounds and their properties are fixed, specifying the density value of sounds fixes the desired diagonal elements of T. This leaves flexibility in determining the MTN parameters, Δ and P, as there are N equations (one for each of the diagonal elements of T) and up to $2N^2$ unknowns. Therefore, we allow P to be chosen by the designer (human or computer). By choosing P, the connecting edges of the network may be defined, and connections between relevant or logically successive sounds may be reinforced with high probability. The desired densities may then be achieved through the necessary values of Δ.

To determine Δ, we first define $F = E\Phi$, where $E \in \mathbb{R}^{N \times N}$, $\Phi \in \mathbb{R}^{N \times N^2}$, and the i^{th} row of E is given by

$$E(i,:) \quad = \quad e_i + q_i(I - Q_i)^{-1}E_i, \tag{12}$$

where Q_i is P with the i^{th} column and row removed, q_i is the i^{th} row of P with $P(i,i)$ removed, e^i is the i^{th} row of the size-N identity matrix, E_i is the identity matrix with the i^{th} row removed, and Φ consists of all zeros except for

$$\Phi(i, i + N * (j-1)) \quad = \quad P(i,j), \tag{13}$$

where i and j are iterated from 1 to N. Finding Δ may then be achieved by solving the quadratic program:

$$\begin{aligned} \text{Minimize} \quad & \|F \cdot \mathbf{vec}(\Delta) - \tau\|_2^2 \\ \text{subject to} \quad & \mathbf{vec}(\Delta) \succeq b \end{aligned}$$

where b is a vector of elements greater than or equal to zero, and $\tau \in \mathbb{R}^N$ is the column vector where the i^{th} element is the value of $T(i,i)$ necessary to achieve the desired density of sound i. The inequality constraint is introduced to allow future extensions where a minimum delay time between certain sounds may be desired. We note that the amount of nontrivial elements of Δ is limited by the edges of the network, and that in some cases the actual set of achieved densities may be the best approximation of densities in a squared error sense.

3.2. Automated Model Design

Using information from the ontological framework and the sounds themselves, we have developed a method of automatically designing an MTN to re-sonify the sound activity of a specified "virtual location" that corresponds to a physical location. Seeking to play the sounds from and relevant to the location, we use our ontological framework to make connections between relevant sounds in the MTN and specify the other parameters such that the expected densities of local sounds are relatively high. By making local sounds dense in the soundscape, they will clearly be heard often, making the available local sounds a key component of the soundscape. As this also implies that the actant process will often travel to local sounds, the creation of edges based on relevancy and importance may aid the actant process in traversing nodes corresponding to sounds relevant to the recorded local sounds. This method is executed as follows.

The edges between vertices are determined by performing a Delaunay triangulation (the dual graph of a Voronoi tessellation) on the sound locations in the previously described two-dimensional MDS. Where a line is drawn between two vertices in the triangulation, edges will be created in both directions; self-connections are not made. The results of Delaunay triangulation on the MDS vary with the placement and clustering of sounds, but it generally connects sounds to those nearby (i.e., sounds deemed relevant by the ontological framework) in the MDS. These connections allow the playback of local sounds to often be preceded and/or succeeded by relevant sounds. The triangulation, however, also makes some connections between sounds considered irrelevant to one another, but the inclusion of such connections can help to ensure that actant processes do not always concentrate near certain nodes when the local sounds are spread in the MDS. Use of Delaunay triangulation also guarantees that every vertex will be connected to at least two other vertices, which can help to prevent repetition.

The desired density of sounds is specified to be inversely related to the distance between the sound's location of recording and the user's virtual location. Currently, we implement this relation as a Gaussian function, referring to the standard deviation as the "listening radius," which sets the size (in surface area) of the region to be explored. The total density of all sounds may be adjusted (so that soundscapes are not overly sparse or dense), perhaps most usefully to a constant value. As described in Section 3.1, specification of the densities determines the values of the transition times, but requires transition probabilities to be provided. The probabilities may be set arbitrarily, but the choice of probability distribution will affect the achievable densities of sounds. Currently, we set the probabilities so that they may further "encourage" the actant process to travel to local sounds. We achieve this by setting the transition probabilities between nodes such that the ratios between the probabilities of edges emanating from a node are equal to the ratios of the desired densities of the nodes toward which they are directed. In practice, we have observed that our current distribution scheme typically provides better actual densities than a uniform distribution.

As this method of soundscape synthesis only requires a virtual location as input when sounds and their corresponding ontological framework are available, it may be applied to various interactions, static or dynamic. Presently, we have created an offline interaction that exactly implements our method of automated design. We have also used this scheme (with sub-optimal calculation of the transition times) in an interactive map, to allow virtual soundwalks, where the network's parameters are periodically updated as the virtual location is changed. This map is also a component of a larger social networking website we are developing that can aid in the collection and tagging of sounds.

4. SOUNDWALKS: AN APPLICATION

We are currently developing a social network website, called "Soundwalks" (http://www.soundwalks.org), to facilitate the collection and re-sonification of sounds and community information.

Through this website, we aim to gather sounds to extensively represent geographic sound activity, and especially where there are too few or too many sounds, utilize user-provided information to determine the importance of individual sounds in the re-sonification of geographic sound activity. On the website, users can upload recorded sounds along with GPS data and provide tags to any sounds. Information, including plots of acoustic features over time, for individual sounds is availabe to view. The site also features an interactive map, similar to systems such as [3, 4, 5, 27], in which the location of sound recordings are marked with icons. With these icons, users can listen to the recordings or retrieve relevant data about them. In addition to allowing users to inspect and playback individual sounds, our map has a "virtual soundwalk" mode that allows users to listen to synthesized soundscapes by moving a token across the map. The following subsections describe the major components of the website.

4.1. Sound Capture and Organization

Presently, sounds in our database have been gathered via mobile sound recording equipment used in conjunction with GPS recording devices. Recordings, which range from short transient events to minute-long soundscapes, consist of both selective recordings and those extracted by automated event segmentation from continuous recordings [19]. An interface allows the uploading of the sounds and GPS data from real soundwalks. After uploading, acoustic feature analysis is performed on the sounds, and they are added to the database and ontological framework. To use existing technology to make the capture and uploading of soundwalks easier and more accessible, we are currently developing a mobile phone application for concurrently recording sounds, GPS data, and any other relevant information (e.g., time).

Once a soundwalk is uploaded and analyzed, users may see a list of sounds from the soundwalk and a small map showing where the sounds were recorded. Each sound may be individually inspected by clicking to navigate to a page that presents all current information about the sound. This includes typical computer-file relevant data (e.g., size, creation time) but additionally displays plots of acoustic feature data and a tag cloud (to which the user may contribute).

4.2. Interaction

An interactive audiovisual display, in the form of a map, is the integral component of our application. This map provides a geographic view of the recordings, displayed as icons at their location of recording on the map. To provide a visual cue of sound content, each icon is colored by mapping its location in the two-dimensional ontological framework MDS to a hue-saturation space. Similarly, all tags on the website are colored by their location in the MDS.

As a user navigates through the map, sound dots within a certain distance threshold, as determined by the map zoom level, will merge together as clusters represented as colored dots. The color and the radius of the cluster are dynamically adjusted by calculating the mean of colors and the number of the child sounds, respectively. When the user clicks on a sound icon, an information window will appear, displaying tags and other information about the recording along with an option to play the sound.

Users may also navigate the map using a virtual soundwalk mode, "scrubbing" a virtual token across the map, creating a virtual soundscape. The soundwalk mode has a variable "listening radius" that may be thought of as the radius of a circle that contains the sounds most expected to be heard. It is effectively the size of the area considered in creating the soundscape. The listening radius may be varied from small to large so as to create soundscapes that range from simulating observable soundscapes at specified locations to providing sonic summaries of large geographic regions. The soundscape is created from an automatically generated MTN (as specified in Section 3), using a single actant process. Periodic updates of the network parameters are made to adapt to the user's movement. The actant process (which is initialized to the sound recorded nearest the virtual location) functions continuously, using the MTN as it is updated. A screenshot of the interactive map (with an open information window) in the virtual soundwalk mode appears in Figure 5.

5. RESULTS

The described system has thus far been informally reviewed by select users and the authors with generally favorable assessments. Our database is presently sparsely populated with sounds recorded in a university environment particularly rich with human activity, traffic (both ground and air), and birds. Using our interaction, we have noted that in exploring areas with few recordings, the inclusion of sounds from elsewhere has provided a soundscape we believe to be indicative of the areas with which we are familiar. The most frequent problem we have observed is the inclusion of keynote sounds (e.g., the beep of a light rail car, or the cheer of a stadium's crowd) in inappropriate areas. Additionally, some sounds are played too frequently. More sounds and users are needed to thoroughly assess how our method of soundscape design can perform across differing communities and with a dense collection of sounds, but present results are promising. Formal listening tests are planned to assess the quality of our synthesized soundscapes, both in comparison to actual recorded soundscapes and in terms of community knowledge.

6. CONCLUSIONS AND FUTURE WORK

We have presented a system for the automated re-sonification of geographic sound activity. This re-sonification takes advantage of an ontological framework we have developed that uses acoustic, semantic, and social information that is largely defined by the community. Additionally, we have developed a website application that implements our system for soundscape synthesis. Listening tests and user studies are needed to fully assess our methodology and application.

Areas of future research include use of time data (to provide sound summaries of different points in time) and possible personalization of soundscapes by tracking user activity or preferences. Note that given our current soundscape design method, if certain sounds are precluded (filtered out by time or other data), the MTN may be easily re-designed. Other possible extensions include application of spatialization/reverb effects to enhance the "sense of place" and separate synthesis techniques for different sound types (keynotes, signals, soundmarks), possibly classified by users. Also, we consider the possibility of using geographic-specific information from other sources (such as directory listings or business review sites) to supplement our evaluation of the activity in a space.

7. REFERENCES

[1] Microsoft Live Labs, "Photosynth," http://livelabs.com/photosynth/.

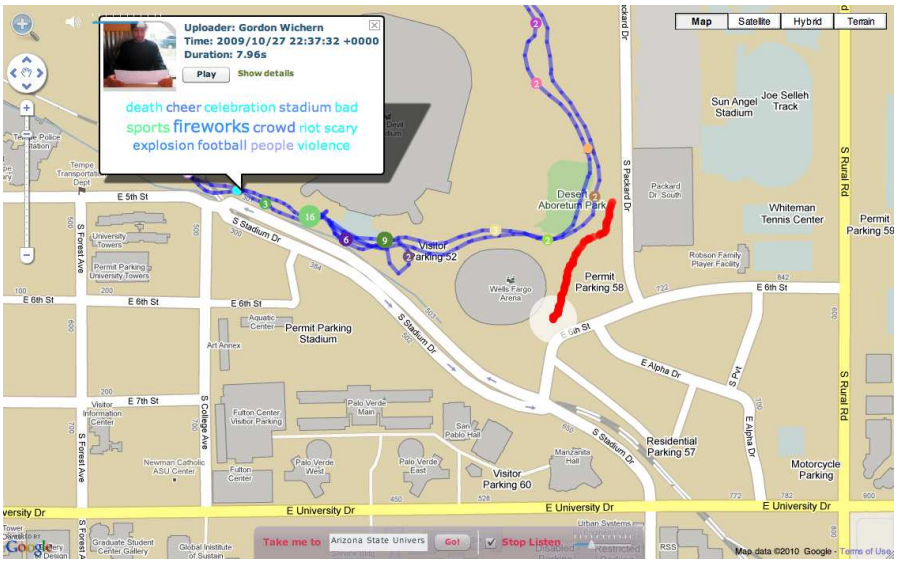


Figure 5: A screenshot of the interactive Soundwalks map in the virtual soundwalk mode.

[2] Google Inc., "Street View," http://maps.google.com/help/maps/streetview/.

[3] Universitat Pompeu Fabra Music Technology Group, "The Freesound Project," http://www.freesound.org/.

[4] Open Sound New Orleans, "Open sound new orleans," http://www.opensoundneworleans.com/core/.

[5] Sound Around You, "Sound around you," http://soundaroundyou.com/.

[6] R. Schafer, *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester, VT: Destiny Books, 1977.

[7] B. Truax, *Acoustic Communication*. Norwood, NJ: Ablex Publishing, 1984.

[8] S. Feld, "Waterfalls of song: an acoustemology of place resounding in bosavi, papua new guinea," in *Senses of Place*, S. Feld and K. Basso, Eds. Santa Fe, NM, USA: School of American Research Press, 1996, pp. 91–136.

[9] A. Misra, P. R. Cook, and G. Wang, "Musical tapestry: Recomposing natural sounds," in *Proceedings of the International Computer Music Conference (ICMC)*, New Orleans, LA, USA, 2006.

[10] P. Cano, L. Fabig, F. Gouyon, M. Koppenberger, A. Loscos, and A. Barbosa, "Semi-automatic ambiance generation," in *Proceedings of the International Conference of Digital Audio Effeccts (DAFx04)*, Naples, Italy, 2004.

[11] S. Serafin, "Sound design to enhance presence in photorealistic virtual reality," in *Proceedings of the 2004 International Conference on Auditory Display*, Sydney, 2004.

[12] A. Valle, M. Schirosa, and V. Lombardo, "A framework for soundscape analysis and re-synthesis," in *Proceedings of the Sound and Music Computing Conference*, Porto, Portugal, 2009.

[13] A. Valle, V. Lombardo, and M. Schirosa, "A graph-based system for the dynamic generation of soundscapes," in *Proceedings of the 15th International Conference on Auditory Display*, Copenhagen, 2009, pp. 217–224.

[14] I. McGregor, A. Crerar, D. Benyon, and C. Macaulay, "Soundfields and soundscapes: Reifying auditory communities," in *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, 2002.

[15] D. Birchfield, N. Mattar, and H. Sundaram, "Design of a generative model for soundscape creation," in *Proceedings of the International Computer Music Conference (ICMC)*, Barcelona, Spain, 2005.

[16] Google Inc., "Corporate Information - Technology Overview," http://www.google.com/corporate/tech.html.

[17] G. Wichern, H. Thornburg, and A. Spanias, "Unifying semantic and content-based approaches for retrieval of environmental sounds," in *Proceedings of the IEEE Wokshop on the Applications of Signal Processing to Audio and Acoustic (WASPAA)*, New Paltz, NY, 2009.

[18] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*, 2nd ed. Cambridge, MA: MIT Press and McGraw-Hill, 2001.

[19] G. Wichern, H. Thornburg, B. Mechtley, A. Fink, A. Spanias, and K. Tu, "Robust multi-feature segmentation and indexing for natural sound environments," in *IEEE CBMI*, Bordeaux, France, July 2007.

[20] G. Wichern, J. Xue, H. Thornburg, and A. Spanias, "Distortion-aware query by example for environmental sounds," in *Proceedings of the IEEE Wokshop on the Applications of Signal Processing to Audio and Acoustic (WASPAA)*, New Paltz, NY, 2007.

[21] B. H. Huang and L. R. Rabiner, "A probabilistic distance measure for hidden Markov models," *AT&T Tech. Journal*, vol. 64, no. 2, pp. 1251–1270, 1985.

[22] T. Pederson, S. Patwardhan, and J. Michelizzi, "WordNet::Similarity - Measuring the Relatedness of Concepts," in *AAAI-04*. Cambridge, MA: AAAI Press, 2004, pp. 1024–1025.

[23] A. Budanitsky and G. Hirst, "Semantic distance in WordNet: an experimental, application-oriented evaluation of five measures." in *Workshop on WordNet and Other Lexical Resources, Second meeting of the North American Chapter of the Association for Computational Linguistics*, Pittburgh, PA, 2001.

[24] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.

[25] J. Kruskal and M. Wish, *Multidimensional Scaling*. Beverly Hills, CA: Sage Publications, 1978.

[26] A. Valle and V. Lombardo, "A two-level method to control granular synthesis," in *Proc. of the 14-th Colloquium on Musical Informatics*, 2003, pp. 136–140.

[27] Wild Sanctuary, Inc., "Wild sanctuary," http://www.wildsanctuary.com/.