Computationally Created Soundscapes with Audio Metaphor

Miles Thorogood and Philippe Pasquier
School of Interactive Art and Technology
Simon Fraser University
Surrey, BC V3T0A3 CANADA
mthorogo@sfu.ca

Abstract

Soundscape composition is the creative practice of processing and combining sound recordings to evoke auditory associations and memories within a listener. We present *Audio Metaphor*, a system for creating novel soundscape compositions. *Audio Metaphor* processes natural language queries derived from Twitter for retrieving semantically linked sound recordings from online user-contributed audio databases. We used a simple natural language processing to create audio file search queries, and we segmented and classified audio files based on general soundscape composition categories. We used our prototype implementation of Audio Metaphor in two performances, seeding the system with keywords of current relevance, and found that the system produced a soundscape that reflected Twitter activity and kept audiences engaged for more than an hour.

1 Introduction

Creativity is a preeminent attribute of the human condition that is being actively explored in artificial intelligence systems aiming at endowing machines with creative behaviours. Artificial creative systems have simulated or been inspired by human creative processes, including, painting, poetry, and music. The aim of these systems is to produce artifacts that humans would judge as creative. Much of the successful research in musical creative systems has focussed on symbolic representations of music, often with corpora of musical scores. Alternatively, non-symbolic forms of music have been little explored in as much detail.

Soundscape composition is a type of non-symbolic music aimed to rouse listeners memories and associations of soundscapes using sound recordings. A soundscape is the audio environment perceived by a person in a given locale at a given moment. A listener brings a soundscape to mind with higher cognitive functions like template matching of the perceived world with known sound environments and deriving meaning from the triggered associations (Botteldooren et al. 2011). People communicate their subjective appraisal of soundscapes using natural language descriptions, revealing the semiotic cues of soundscape experiences (Dubois and Guastavino 2006).

Soundscape composition is the creative practice of processing and combining sound recordings to evoke auditory associations and memories within a listener. It is positioned along a continuum with concrete music that uses found sound recordings, and electro-acoustic music that uses more abstracted types of sounds. Central to soundscape composition, is processing sound recordings. There are a range of approaches to using sound recordings. One approach is to portray a realistic place and time by using untreated audio recordings, or, recordings with only minor editing (such as cross-fades). Another is to evoke imaginary circumstances by applying more intensive processing. In some cases, these manufactured sound environments appear imaginary, by the combination of largely untreated with more highly processed sound recordings. For example, the soundscape composition *Island*, by Canadian composer Barry Truax (Truax 2009), adds a mysterious quality to a recognizable sound environment by contrasting clearly discernible wave sounds against less-recognizable background drone and texture sounds.

Soundscape composition requires many decisions about selecting and cutting audio recordings and their artistic combination. These processes become exceedingly time consuming for people when large amounts of audio data are available, as is now the case with online databases. As such, different generative soundscape composition systems have automated many sub-procedures of the composition process, but we have not found any systems in the literature to date that use natural language processing for generative soundscape composition. Likewise, automatic audio segmentation for soundscape composition specific categories is an area not yet explored.

The system described here searches online for the most recent Twitter posts about a small set of themes. Twitter provides an accessible platform for millions of discussions and shared experiences through short text-based posts (Becker, Naaman, and Gravano 2010). In our research, audio file search queries are generated from natural language queries derived from Twitter. However, these requests could be a memory described by a user, a phrase from a book, or a section of a research paper.

Audio Metaphor accepts a natural language query (NLQ), which is made into audio file search queries by our algorithm. The system searches online for audio files semantically related to word features in the NLQ. The resulting audio file recommendations are classified and segmented based

upon the soundscape categories *background*, *foreground*, and *background with foreground*. A composition engine autonomously processes and combines segmented audio files.

The title of *Audio Metaphor* refers to the idea that audio representations of NL queries that the system generates may not have literal associations. Although, in some cases, an object referenced in the NL query may have a direct referential sound such as with "raining outside" that results in a type of *audio analogy*. However, an example that is not as direct such as, "A brooding thought struck me down" has no such direct referent to an object in the world. In this latter case, *Audio Metaphor* would create a composition by processing sound recordings that have some semantic relationship with words in the NL query. For example, the sound of a storm and the percussive striking of an object are the types of sounds that would be processed in this case.

Margret A. Boden actively proposes types of creativity being synthesized by computational means (Boden 1998). She states, that *combinatorial* type creativity "involves novel (improbable) combinations of familiar ideas ... wherein newly associated ideas share some inherent conceptual structure." The artificial creative system here uses semantic inference driven by NLQs as a way to frame the soundscape composition and make use of semantic structures inherent in crowdsourced systems. Further to this, the system associates words with sound recordings for combining into novel representations of texts. For this reason, the system is considered to exhibit *combinatorial* creative behaviour.

Our contribution is a creative and autonomous soundscape composition system with a novel method of generating compositions from natural language input and crowd-sourced sound recordings. Furthermore, we present a method of audio file segmentation based on soundscape categories, and a soundscape composition engine that contrasts sound recording segments with different levels of processing.

We outline our research in the design of an autonomous soundscape composition system called *Audio Metaphor*. In the next section, we show the related works in the domains of soundscape studies and generative soundscape composition. We go on to describe the system architecture, including natural language processing, classification and segmentation, and the soundscape composition engine. The system is then disused in terms of a number of performances and presentations. We conclude with our ideas for future work.

2 Related Work

Birchfield, Mattar, and Sundaram (2005) describe a system that uses an adaptive user model for context-aware soundscape composition. In their work, the system has a small set of hand-selected and hand-labelled audio recordings that were autonomously mixed together with minimal processing. Similarly, Eigenfeldt and Pasquier (2011) employ a set of hand-selected and hand-labelled environmental sound recordings for the retrieval of sounds from a database by autonomous software agents. In their work, agents analyze audio when selecting sounds to mix based on low-level audio features. In both cases, listening and searching for selecting audio files is very time consuming.

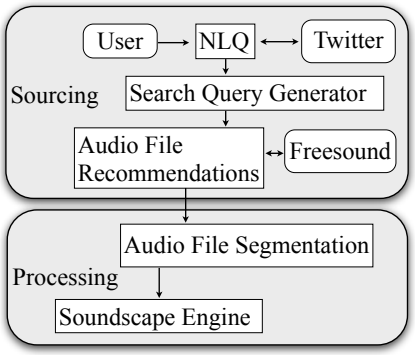

Figure 1: *Audio Metaphor* system architecture overview.

A different approach to selecting and labelling sound recordings is to take advantage of collaborative tagging of online user-contributed collections of sound recordings. This is a crowdsourcing process where a body of tags is produced collaboratively by human users connecting terms to documents (Halpin, Robu, and Shepherd 2007). In online environments, collaborative tags are part of a shared language made manifest by users (Marlow et al. 2006). Online audio repositories such as pdSounds (Mobius 2009) and Freesound (Akkermans et al. 2011) demonstrate collaborative tagging systems applied to sound recordings.

A system that uses collaborative tags to retrieve sound recordings is described by Janer, Roma, and Kersten (2011). In their work, a user defines a soundscape composition by entering locations on a map that has sounds tags associated with various locations. As the user navigates the map, a soundscape is produced. In related research, the locations on a map are used as a composition environment (Finney and Janer 2010). Their compositions use hand-selected sounds, which are placed in close and far proximity based upon semantic identifiers derived from tags.

3 System Architecture

Audio Metaphor creates unique soundscape compositions that represent the words in an NLQ using a series of processes as follows:

- Receive a NLQ from a user, or Twitter;
- Transforms a NLQ into audio file search queries;
- Search online for audio file recommendations;
- Segment audio files into soundscape regions;
- Process and combine audio segments for soundscape composition.

In the *Audio Metaphor* system, these processes are handled by sequentially as is shown in Figure 1. [1]

[1] A modular approach was taken for the system design. Accordingly, the system is flexible to be used for separate objectives, including, making audio file recommendations to a user from an NLQ, and deriving a corpus of audio segments.

rainy autumn day vancouver
rainy autumn day
autumn day vancouver
rainy autumn
autumn day
day vancouver
rainy
autumn
day
vancouver

Table 1: All sub-lists generated from a word-feature list from the query "On a rainy autumn day in Vancouver'.

3.1 Audio File Retrieval Using Natural Language Processing

The audio file recommendation module creates audio file search queries given a natural language request and a maximum number of audio file recommendations for each search.

The Twitter web API (Twitter API) is used to retrieve the 10 most recent posts related to a theme to find current associations. The longest of these posts is then used as a natural language query. To generate audio file search queries, a list of word features is extracted from the input text and generates a queue of all unique sublists. These sublists are used as search queries, starting with the longest first. The aim of the algorithm is to minimize the number of audio files returned and still represent all the word features in the list. When a search query returns a positive result, all remaining queries that contain any of the successful word features are removed from the queue.

To extract the word features from the natural language query, we use essentially the same method as that proposed by Thorogood, Pasquier, and Eigenfeldt (2012), but with some modifications. The algorithm first removes common words listed in the Oxford English Dictionary Corpus, leaving only nouns, verbs, and adjectives. Words are kept in order and treated as a list. For example, with the word feature list from the natural language query "The angry dog bit the crying man," "angry dog bit crying man," is more valid than "angry man bit crying dog."

The algorithm for generating audio file queries essentially extracts all the sublists from the NLQ that have a length greater than or equal to 1. For example, a simple request such as "On a rainy autumn day in Vancouver" is first processed to extract the word feature list: rainy, autumn, day, vancouver. After that, sub-lists are generated as shown in Table 1.

Audio Metaphor accesses the Freesound audio repository for audio files with the Freesound API. Freesound is an online collaborative database with over 120,000 audio clips. The indexed data includes user-entered descriptions and tags. The content of the audio file is inferred from user-contributed commentary and social tags. Although there is no explicit user rating of audio files, a download counter for each file provides a measure of its popularity, and search results are presented by descending popularity count.

The sublists are used to search online for semantically related audio files using an exclusive keyword search. Sublists are used in the order created, from largest to smallest. A search is considered successful when it returns one or more recommendations. Additionally, the algorithm optimizes audio file recommendations by ignoring future sublists that contain word features from a previously successful search. The most favourable result is a recommendation for the longest sub-list, with the worst case being no recommendations. In practice, the worst case is, typically, a recommendation for each singleton word feature.

For each query, the URLs of the recommendations are logged in a separate list. The list is constrained to a number specified at the system startup. Furthermore, if a list has less than the number of files requested it is considered sparsely populated and no further modification made to its items. For example, if the maximum number of recommendations specified for each query is five, and there are two queries where one returns nine recommendations and the other three, the longer list will be constrained to five, and the empty items of the second list are ignored.

The separate lists of audio file recommendations are then presented to the audio segmentation module.

3.2 Audio File Classification and Segmentation

Audio segmentation is an essential preprocessing step in many audio applications (Foote 2000). In soundscape composition, a composer will choose background and foreground sound regions to combine into new soundscapes.

Background and foreground sounds are general categories that refer to a signal's perceptual class. Background sounds seem to come from farther away than foreground sounds or occur often enough to belong to the aggregate of all sounds that make up the background texture of a soundscape. This is synonymous with a ubiquitous sound (Augoyard and Torgue 2006): a sound that is diffuse, omnidirectional, constant, and prone to sound absorption and reflection factors having an overall effect on the quality of the sound. Urban drones and the purring of machines are two examples of ubiquitous or background sound. Conversely, foreground sounds are typically heard standing out clearly against the background. At any moment in a sound recording, there may be either background sound, foreground sound, or a combination of both.

Segmenting an audio file is a process of listening to the recording for salient features and *cutting* regions for later use. To automate this process, we have designed an algorithm to classify segments of an audio file and concatenate neighbouring segments with the same label. An established technique for classification of an audio recording is to use a supervised machine learning algorithm trained with examples of classified recordings.

3.3 Audio Features Used for Segmentation

The classifier models the generic soundscape categories *background*, *foreground*, and *background with foreground*. We use a vector of the low-level audio features total-loudness, and the first three mel-frequency cepstral coefficients (MFCC). These features reflect the behaviour of the human auditory system, which is an important aspect of

soundscape studies. They are extracted at a frame-level from an audio signal with a window of 23 ms and a step size of 11.5 ms using the Yaafe audio feature extraction software package (Mathieu et al. 2010).

MFCC audio features represent the spectral characteristics of a sound by a small number of coefficients calculated by the logarithm of the magnitude of a triangular filter bank. We use an implementation of MFCC that builds a logarithmically spaced filter bank according to 40 coefficients mapped along the perceptual Mel-scale by:

$$M(f) = 1127 \log \left(1 + \frac{f}{700}\right) \quad (1)$$

where f is the frequency in Hz.

Total loudness is the characteristic of a sound associated with the sensation of intensity. The human auditory system affects the perception of intensity of different frequencies. One model of loudness (Zwicker 1961) takes into account the disparity of loudness at different frequencies along the Bark scale, which corresponds to the first 24 critical bands of hearing. Bands near human speech frequencies have a lower threshold than those of low and high frequencies. The conversion from a frequency in Hz f to the equivalent frequency in the Bark scale B is calculated with the following formula (Traunmuller 1990).

$$B(f) = 13 \arctan(0.00076f) + 3.5 \arctan\left(\frac{f}{7500}\right)^2 \quad (2)$$

Where f is the frequency in Hz. A specific loudness is the loudness calculated at each Bark band; the total loudness is the sum of individual specific loudnesses over all bands. Because a soundscape is perceived by a human not at the sample level, but over longer time periods, we use a so called *bag of frames approach* (Aucouturier and Defreville 2007) to account for longer signal durations. Essentially, this kind of approach considers frames that represent a signal have possibly different values, and the density distribution of frames provides a more effective representation than a singular frame. Statistical methods, such as the mean and standard deviation of features, recapitulate the texture of an audio signal, and provides a more effective representation than a single frame.

In our research, audio segments are represented with an eight-dimensional feature vector of the means and standard deviations from the total loudness and the first 3 MFCC. The mean and standard deviation of the feature vector models the *background*, *foreground*, and *background with foreground* soundscape categories well. For example, sounds distant from the listener and considered background sound will typically have a smaller mean total loudness. Sounds that occur often enough will have a smaller standard deviation of those in foreground listening. MFCC takes into account the spectrum of the sound affected by its source placement in the environment.

3.4 Supervised Classifier Used for Segmentation

We used a Support Vector Machine classifier (SVM) to classify audio segments. SVMs have been used in environmental sound classification problems, and consistently demonstrated good classification accuracy. A SVM is a non-probabilistic classifier that learns optimal separating hyperplanes in a higher dimensional space from the input. Typically, classification problems present non-linearly separable data that can be mapped to a higher-dimensional space with a kernel function. We use the C-support vector classification (C-SVC) algorithm shown by Chang and Lin (2011). This algorithm uses a radial basis function as a kernel, which is suited to a vector with a small number of features and takes into account the relation between class labels and attributes being non-linear.

Training Corpus The classifier was trained using feature vectors from a pre-labelled corpus of audio segments. The training corpus consists of 30 segments between 2 and 7 seconds long. Audio segments were labelled from a consensus vote by human subjects in an audio segment classification study. The study was conducted online through a web browser. Audio was played to participants using an HTML5 audio player object. This player allowed participants to repeatedly listen to a segment. Depending on the browser software, the audio format of segments was either MP3 at 196 kps, or Vorbis at an equivalent bit rate. Participants selected a category from a set of radio buttons and each selection was confirmed when the participant pressed a button to listen to the next segment.

There were 15 unique participants in the study group from Canada and the United States. Before the study started, an example for each of the categories, background, foreground, and background with foreground, was played, and a short description of the categories was displayed. Participants were asked to use headphones or audio monitors to listen to segments. Each participant was asked to listen to the randomly ordered soundscape corpus. On completing the study, the participant's classification results were uploaded into a database for analysis.

The results of the study were used to label the recordings by a majority vote. Figure 2 shows the results of the vote. Results of the vote gave the labelling to the recordings. There are a total of 10 recordings for each of the categories.

A quantitative analysis of the voter results shows the average agreement of recordings for each category as follows: background 84.6% (SD=18.6%); foreground 77.0% (SD=10.4%), and; background with foreground 76.2% (SD=13.4%). The overall agreement was shown to be 79.3% (SD=4.6%).

Classifier Evaluation We evaluated the classifier, using the training corpus, with a 10-fold cross validation. The results summary is shown in Table 2. The classifier achieved an overall sample accuracy of 80%, which shows that the classifier was human competitive against the overall human agreement statistic of 79.3%.

The kappa statistic is a chance-corrected measure showing the accuracy of prediction among each k-fold model. A kappa score of 0 means the classifier is performing only as well as chance; 1 implies a perfect agreement; and a kappa score of .7 is generally considered satisfactory. The kappa score of .7 in the results shows a good classification accuracy was achieved using the described method.

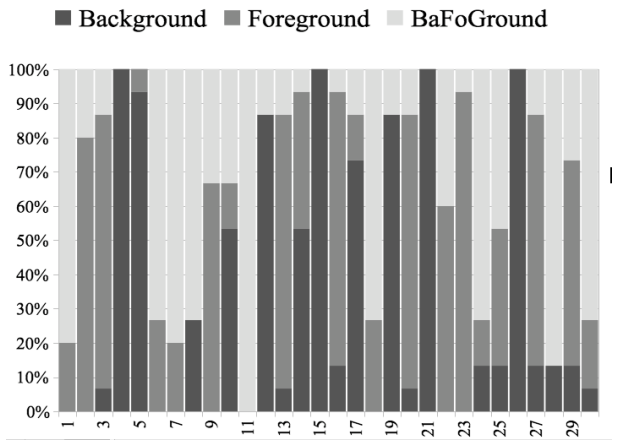

Figure 2: Audio classification vote results from human participants for 30 sound recordings with three categories: Background, Foreground, and Background with Foreground (BaForound) sound.

Table 2: Summary of SVM classifier with the mean and standard deviation for features total loudness and 3 MFCC.

Correctly classified instances	24	80%
Incorrectly classified instances	6	20%
Kappa statistic	0.7	

These performance measures are reflected by the confusion matrix in Table 3. All 10 of the audio segments labelled "background" from the study were classified correctly. The remaining audio segments, labelled "foreground" and "background with foreground," were correctly classified 7 out of 10 times, with the highest level of confusion between these latter categories.

3.5 Background-Foreground Segmentation

In our segmentation method, we use a 500 ms sliding analysis window with a hop size of 250 ms. We found that for our application an analysis window of this length provided reasonable information for the bag of frames approach and ran with satisfactory computation time. The resulting feature vector is classified and labelled as belonging to one of the three categories. In order to create labelled regions of more than one window, neighbouring windows with the same label are concatenated and the start and end time of the new window are logged.

To demonstrate the segmentation algorithm, we used a 9 second audio file containing a linear combination of background, foreground, and background with foreground regions. Figure 3. shows the ground truth with the solid black line, and algorithm segmentation of the audio file with background, foreground, and background with foreground labelled regions applied. We use the SuperCollider3 software package for visualizing the segmented waveform sc3. This example shows concatenated segments labelled as regions. One of the background with foreground segments was misclassified resulting in a slightly longer foreground region than the ground truth classification.

Table 3: Confusion matrix of SVM classifier for the categories background (BG), foreground (FG), and background with foreground (BgFg).

Bg	Fg	BgFg	
10	0	0	Bg
0	7	3	Fg
1	2	7	BgFg

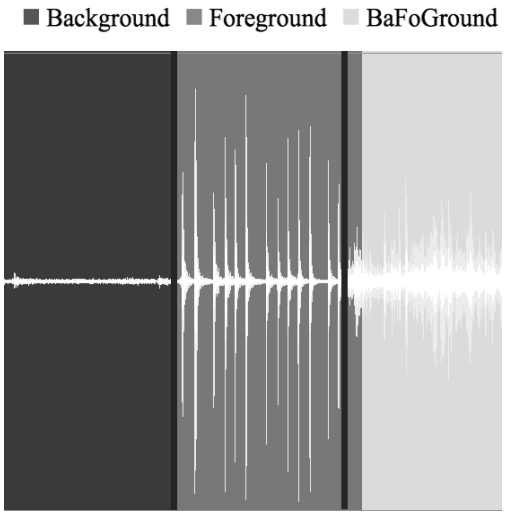

Figure 3: Segmentation of the audio file with ground-truth regions (black line) and segmented regions Background (dark-grey), Foreground (mid-grey), and Background with Foreground (light-grey).

The audio files and the accompanying segmentation data are then presented to the composition module.

3.6 Composition

The composition module creates a layered two-channel soundscape composition by processing and combining classified audio segments. Each layer in the composition consists of processed background, foreground, and background with foreground sound recordings. Moreover, an agent-based model is used in conjunction with a heuristic in order to handle different sound recordings and mimic the decisions of a human composer. Specifically, we based this heuristic from production notes for the soundscape composition *Island*, by Canadian composer Barry Truax. In these production notes, Truax gives detailed information on how sound recordings are effected, and the temporal arrangement of sounds.

In our modelling of these processes, we chose to use the first page of the production notes, which corresponds to around 2 minutes of the composition. Furthermore, we framed the model to comply with the protocol of the seg-

mentation labels and aesthetic evaluations by the authors. A summary of the model is as follows:

- Regions labelled *background* are played sequentially in the order presented by the segmentation. They are processed to form a dramatic textured background. This processing is carried out by first playing the region at 10% of its original speed and applying a stereo time domain granular pitch shifter with ratios 1:0.5 (down an octave) and 1:0.667 (down a 5th). We added a Freeverb reverb (Smith 2010) with a room size of 0.25 to give the texture a more spacious quality. A low pass filter with a cutoff frequency at 800 Hz is used to obscure any persistent high end detail. Finally, a slow spatialization is applied in the stereo field at a rate of 0.1 Hz.
- Regions labelled *foreground* are chosen from the foreground pool by a roll of the dice. They are played individually, separated by a period proportional to the duration of the current region played $t = d^{.75} + d + C$, where t is the time between playing the next region, d is the duration of the current region, and C is a constant controlling the minimum duration between regions. In order to separate them from the background texture, foreground regions are processed by applying a band pass filter with a resonant frequency 2,000 Hz and high Q value of 0.5. Finally, a moderate spatialization is applied in the stereo field at a rate of .125 Hz.
- Regions labelled *background with foreground* are slowly faded in and out to evoke a mysterious quality to the soundscape. They are chosen from the pool of regions by a roll of the dice and are played for an arbitrarily chosen duration of between 10 and 20 seconds. Regions with a length less than the chosen duration are looped. In order to achieve a separation from the background texture and foreground sounds, regions are processed by applying a band pass filter with a resonant frequency 8,000 Hz and high Q value of 0.1. The addition of a Freeverb reverb with a room size of 0.125 and a relatively fast spatialization at a rate of 1 Hz was used to further add to the mysterious quality of the sound.

This composition model is deployed individually by each of agents of the system, who are responsible for processing a different audio file. An agents decisions are, choosing labelled regions of an audio recording, processing and combining them in a layered soundscape composition according to the composition model.

Because of the potentially large number of audio files available to the system, and in order to limit the acoustic density of a composition, a maximum number of agents are specified on system start-up. If there are more audio file results than there are agents to handle them, the extra results are ignored. Equally, if the number of results is smaller then the number of agents, agents without tasks are temporarily ignored.

An agent uses the region labels of the audio file to decide which region to process. An audio file may have a number of labelled regions. If there is no region of a type then that type is ignored. The agent can play one of each types of region simultaneously.

4 Qualitative Results

Audio Metaphor has been used in performance environments. In one case, the system was seeded with the words "nature," "landscape," and "environment." There were roughly 150 people in the audience. They were told that the system was responding to live Twitter posts and shown the console output of the search results. During the performance, there was an earthquake off the coast of British Columbia, Canada, and the current Twitter posts focused on news of the earthquake. *Audio Metaphor* used these as natural language requests, searched online for sound recordings related to earthquakes, and created a soundscape composition. The sound recordings processed by the system included an earthquake warning announcement, the sound of alarms, and a background texture of heavy destruction. The audience reacted by checking to see if this event was indeed real. This illustrated how the semantic space of the soundscape composition effectively maps to the concepts of a natural language request.

In a separate performance, *Audio Metaphor* was presented to a small group of artists and academics. This took place during the height of the 2012 conflict in Syria, and the system was seeded with the words "Syria," "Egypt," and "conflict." The soundscape composition presented segments of spoken word, traditional instruments, and other sounds. The audience listened to the composition for over an hour without losing its engagement with the listening experience. One comment was, "It was really good, and we didn't get bored." The sounds held peoples' attention because they were linked to current events, and the processing of sound recordings added to the interest of the composition.

Because the composition model deployed in *Audio Metaphor* is based of a relatively short section of a composition, there was not a great deal of variation in the processing of sound recordings. The fact that people were engaged for such long periods of time suggests that other factors contributed to the novel stimulus. Our nascent hypothesis is that the dynamic audio signal of recordings, in addition to the processing of audio files contributed to listeners ongoing engagement. [2]

5 Conclusions and Future Work

We describe a soundscape composition engine that chooses audio segments using natural language queries, segments and classifies the resulting files, processes them, and combines them into a soundscape composition at interactive speeds. This implementation uses current Twitter posts as natural language queries to generate search queries and retrieves audio files that are semantically linked to queries from the Freesound audio repository.

The ability of *Audio Metaphor* to respond to current events was shown to be a strong point in audience engagement. The presence of signifier sounds evoked listeners' associations of concepts. Listener engagement was further reinforced through the artistic processing and combination of sound recordings.

[2] Sound examples of *Audio Metaphor* using the composition engine can be found at http://www.audiometaphor.ca/aume

Audio Metaphor can be used to help sound artists and autonomous systems retrieve and cut sound field recordings from online audio repositories. Although, its primary function, as we have demonstrated, is autonomous machine generated soundscapes for performance environments and installations. In the future, we will evaluate people's response to these compositions by distributing them to user-contributed music repositories and analyzing user comments. These comments can then be used to inform the *Audio Metaphor* soundscape composition engine.

Although the system generates engaging and novel soundscape compositions, the composition structure is tightly regulated by the handling of background and foreground segments. In future work, we aim toward equipping our system with the ability to evaluate its audio output, in order to make more in-depth composition decisions. By developing these methods, *Audio Metaphor* will be not only be capable of processing audio files to create novel compositions, but, additionally, be able to respond to the compositions it has made.

6 Acknowledgments

This research was funded by a grant from the Natural Sciences and Engineering Research Council of Canada. The authors would also like to thank Barry Truax for his composition and production documentation.

References

Akkermans, V.; Font, F.; Funollet, J.; de Jong, B.; Roma, G.; Togias, S.; and Serra, X. 2011. Freesound 2: An Improved Platform for Sharing Audio Clips. In *International Society for Music Information Retrieval Conference*.

Aucouturier, J.-J., and Defreville, B. 2007. Sounds like a park: A computational technique to recognize soundscapes holistically, without source identification. *19th International Congress on Acoustics*.

Augoyard, J., and Torgue, H. 2006. *Sonic Experience: A Guide to Everyday Sounds*. McGill-Queen's University Press.

Becker, H.; Naaman, M.; and Gravano, L. 2010. Learning similarity metrics for event identification in social media. In *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, 291–300. New York, NY, USA: ACM.

Birchfield, D.; Mattar, N.; and Sundaram, H. 2005. Design of a generative model for soundscape creation. In *International Computer Music Conference*.

Boden, M. A. 1998. Creativity and artificial intelligence. *Artificial Intelligence* 103(1–2):347 – 356.

Botteldooren, D.; Lavandier, C.; Preis, A.; Dubois, D.; Aspuru, I.; Guastavino, C.; Brown, L.; Nilsson, M.; and Andringa, T. C. 2011. Understanding urban and natural soundscapes. In *Forum Acusticum*, 2047–2052. European Accoustics Association (EAA).

Chang, C.-C., and Lin, C.-J. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2:27:1–27:27. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

Dubois, D., and Guastavino, C. 2006. *In search for soundscape indicators : Physical descriptions of semantic categories*.

Eigenfeldt, A., and Pasquier, P. 2011. Negotiated content: Generative soundscape composition by autonomous musical agents in coming together: Freesound. In *Proceedings of the Second International Conference on Computational Creativity*, 27–32. México City, México: ICCC.

Finney, N., and Janer, J. 2010. Soundscape Generation for Virtual Environments using Community-Provided Audio Databases. In *W3C Workshop: Augmented Reality on the Web*.

Foote, J. 2000. Automatic audio segmentation using a measure of audio novelty. In *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*, volume 1, 452 –455 vol.1.

Halpin, H.; Robu, V.; and Shepherd, H. 2007. The complex dynamics of collaborative tagging. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, 211–220. New York, NY, USA: ACM.

Janer, J.; Roma, G.; and Kersten, S. 2011. Authoring augmented soundscapes with user-contributed content. In *ISMAR Workshop on Authoring Solutions for Augmented Reality*.

Marlow, C.; Naaman, M.; Boyd, D.; and Davis, M. 2006. Ht06, tagging paper, taxonomy, flickr, academic article, to read. In *Proceedings of the seventeenth conference on Hypertext and hypermedia*, HYPERTEXT '06, 31–40. New York, NY, USA: ACM.

Mathieu, B.; Essid, S.; Fillon, T.; J.Prado; and G.Richard. 2010. YAAFE, an Easy to Use and Efficient Audio Feature Extraction Software. In *Proceedings of the 2010 International Society for Music Information Retrieval Conference (ISMIR)*. Utrecht, Netherlands: ISMIR.

Mobius, S. 2009. pdSounds. Available online at http://www.pdsounds.org/; visited on April 12th 2012.

Smith, J. O. 2010. *Physical Audio Signal Processing*. W3K Publishing. online book.

Thorogood, M.; Pasquier, P.; and Eigenfeldt, A. 2012. Audio metaphor: Audio information retrieval for soundscape composition. In *Proceedings of the 6th Sound and Music Computing Conference*.

Traunmuller, H. 1990. Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America* 88(1):97–100.

Truax, B. 2009. Island. In Soundscape Composition DVD. DVD-ROM (CSR-DVD 0901). Cambridge Street Publishing.

Twitter API. Available online at https://dev.twitter.com/docs/; visited on April 12th 2012.

Zwicker, E. 1961. Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen). *The Journal of the Acoustical Society of America* 33(2):248.