A Web Search Engine for Sound Effects

Stephen V. Rice[1] and Stephen M. Bailey[2]

[1] The University of Mississippi, University, MS, 38677, USA
rice@cs.olemiss.edu

[2] Comparisonics Corporation, Grass Valley, CA, 95945, USA
sbailey@comparisonics.com

ABSTRACT

FindSounds is the first Web search engine for sound effects. Queries are processed using a selective index of Web audio files that includes sound effects and musical instrument samples but excludes song and speech recordings. A text search retrieves audio files based on how they are labelled, and a "sounds-like" search locates audio files based on sound similarity. Each month FindSounds processes more than 1.5 million queries for more than 150,000 Internet users.

1. INTRODUCTION

Although there are millions of Web sites hosting text, image, audio, and/or video files, the number of Web search engines available for searching these files is small. Figure 1 lists ten Web search engines and the types of files they can retrieve. As of July 2005, only four of these search audio files on the Web: AllTheWeb, AltaVista, FindSounds, and Singingfish.

Audio recordings can be divided into three categories: (1) song recordings, (2) speech recordings, and (3) sound effects. AllTheWeb and AltaVista index all types of audio files on the Web, whereas Singingfish focuses on streaming audio. FindSounds is focused on sound effects.

FindSounds (www.FindSounds.com) was developed by the authors of this article. Since its debut on August 1, 2000, it has processed more than 50 million sound searches. Each month it handles more than 1.5 million queries from more than 150,000 unique visitors.

AllTheWeb	text, image, audio, video
AltaVista	text, image, audio, video
Ditto	image
FindSounds	audio
Google	text, image
MSN Search	text, image
Picsearch	image
Singingfish	audio, video
Teoma	text
Yahoo!	text, image, video

Figure 1 Web Search Engines

FindSounds appeals to Internet users of all ages and is especially valuable to sound designers, musicians, filmmakers, videographers, animators, and game developers. Like any content found on the Web, audio files may contain copyrighted material and it is the user's obligation to obtain copyright clearance if required for the intended use.

Suppose an Internet user wishes to locate elephant sounds on the Web. If audio files are indexed indiscriminately, a search for "elephant" may return more songs about elephants (e.g., Henry Mancini's *Baby Elephant Walk*), and recordings of people speaking about elephants, than elephant sounds. The FindSounds index is highly selective, containing only sound effects, so that such a query returns only elephant sounds.

The category of sound effects is defined broadly here. Although song and speech recordings are excluded from the FindSounds index, musical instrument samples are included, as well as non-speech utterances of the human voice, such as coughs, grunts, and screams. The index includes a large variety of sounds from animals, birds, and insects; office and household environments; and vehicles, explosions, guns, sirens, whistles, etc. The development of the FindSounds index is described in Section 2.

Queries are processed using the FindSounds index. Sounds are difficult to describe in words. The limitations of searching for sounds based on their text descriptions have motivated the development of advanced algorithms for searching recordings based on how they sound, rather than how they have been labelled. FindSounds is the first Web search engine to incorporate "content-based" audio retrieval. An overview of FindSounds query processing is given in Section 3.

2.   INDEX CREATION

A "spider" program "crawls" the Web in search of digital audio files in Wave, AU, or AIFF format, which are the most popular formats for sound effects on Windows, Unix, and Macintosh systems, respectively. Each file is downloaded and analyzed by the program to determine if it is a candidate for inclusion in the FindSounds index. The file must contain a recording of duration between 0.05 and 10 seconds and sample rate between 8 and 96 kHz. The maximum amplitude is checked to exclude recordings that are too quiet. The

DC offset is computed to eliminate poor-quality recordings with excessive offsets.

Only 6% of the audio files located by the spider become candidates for the FindSounds index. In a streamlined process, each candidate is played to a human auditioner who decides whether to accept or reject it. If the recording contains a sequence of three or more different notes or chords, or includes one or more spoken words, then it is considered a song or speech recording and is rejected. Any recording deemed obscene is also rejected, so that FindSounds is safe for children to use.

About 15% of the candidates are accepted by this auditioning process. However, before they are incorporated into the FindSounds index, a human cataloger listens to each recording and attempts to enter a text description of the recording for retrieval purposes. If the sounds defy description, then a description is not entered, yet the recording is still retrievable by a "sounds-like" search. About 58% of the audio files in the FindSounds index receive text descriptions.

Duplicate files are identified automatically in the indexing process. As many as 367 identical copies of a recording have been located. Copies of a previously-rejected recording need not be re-auditioned. For each accepted recording, the URL of only one copy resides in the index, so that users will not receive many links to identical files in a query result. However, the URLs of copies are saved in an internal database so that if one copy goes offline (becomes unavailable), another copy can be substituted. Note that the FindSounds index contains only links to audio files, not the audio files themselves.

To date, the FindSounds spider has located approximately 10 million audio files on the Web. Of these, about 600,000 candidates have been identified and auditioned, from which about 100,000 have been accepted for the FindSounds index. However, because files on the Web become inaccessible over time, the current number of indexed recordings is about 50,000.

The indexed recordings are mostly mono (89% versus 11% stereo) and in Wave format (88% versus 7% AU and 5% AIFF). The average file size is 98 kilobytes. The average duration is 2.75 seconds; see Figure 2 for the distribution of durations. The average sample rate is 22 kHz; see Figure 3 for the distribution of sample rates.
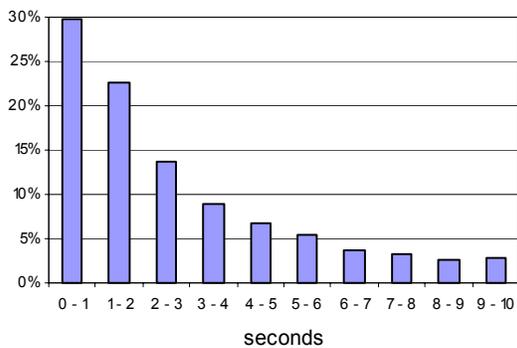
Figure 2 Distribution of Audio File Durations
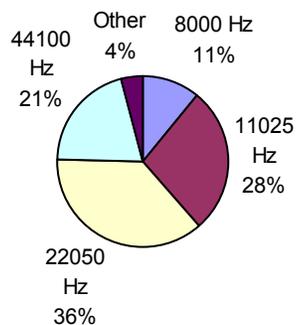
Figure 3 Distribution of Audio File Sample Rates

3.   QUERY PROCESSING

The FindSounds index is placed in computer memory for fast access by the FindSounds query processor. Upon receiving a query, the query processor uses the index to determine the search results. From these results, an HTML page is constructed containing a list of hits, which is returned to the user who posed the query.

There are three types of queries that may be posed:

(1) a text search;

(2) a "sounds-like" search; and

(3) a "combination" search, which is both a text search and a sounds-like search.

For each type of query, the user may restrict the search results using the "metadata controls." With these controls, the user can limit the results to particular file formats and number of channels, and specify the minimum resolution and sample rate. Users with slow Internet connections can limit the size of audio files.

A text search is posed by entering a word or phrase into a search box and clicking on the Search button, or by clicking on one of the "keyword links" on the "Sound Types" page. Using an inverted keyword index, the FindSounds query processor efficiently locates audio files in the FindSounds index that have been labelled using one or more of the submitted words. Up to 200 hits are retrieved and displayed to the user, ten to a page.

Figure 4 illustrates a search box in which a user has entered "owl." The metadata controls beneath the search box show their default settings. Clicking on the Search button retrieves 200 hits labelled "owl." Figure 5 displays the first five of these hits.

Figure 4 Search Box and Metadata Controls

The user can click on the URL of a hit, or the Play icon beside it, to download and play the audio file. The default audio player will be used, typically Windows Media Player on computers running Microsoft Windows.

The text description of a hit appears in bold lettering beneath the URL. It is followed by a line containing the file size in kilobytes, number of channels (mono or stereo), bit resolution, sample rate, and duration in seconds. Clicking on the "show page" link displays a Web page that refers to this audio file, which may

contain copyright information. The "e-mail this sound" link can be used to e-mail the audio file URL.

For each hit, a colored waveform display provides a visual representation of the sounds in the file. The waveform display is color coded to represent the frequency content. Sounds dominated by high frequencies are colored with red shades; mid-range sounds are shaded green or blue; and dark colors convey low (bass) tones. Users learn to "read" the waveform display, that is, they obtain an impression of how a recording will sound by inspecting the display, which helps them decide which audio files to download and audition. More information on the colored waveform display is provided in a companion paper [1].

Figure 5 Text Search Results

The sound of each audio file in the FindSounds index is characterized by a 16-byte "signature," which is stored in the index. An algorithm compares any two signatures and computes a similarity score ranging from 0% (least similar) to 100% (most similar, i.e., identical). This score is an estimate of the perceived similarity of the two recordings from which the signatures were derived.

In a sounds-like search, any audio file may be used as an example sound, and the query processor returns a list of similar-sounding files in order of decreasing similarity to the example. The matches are determined without regard to how the sounds are labelled.

Clicking on the Sounds-Like Search icon ⌖ beside the first hit in Figure 5 causes the 200 most similar sounds in the FindSounds index to be retrieved (see Figure 6). Note that the owl recording matched two unlabelled sounds (hits 1 and 2) with similarity scores of 89% and 85%, respectively. Similar sounds may come from any source, such as an eagle or telephone. Note the similar colors in the waveform displays, reflecting the similar frequency content.

Figure 6 Sounds-Like Search Results

In a combination search, the results of a sounds-like search are restricted to show only the matches that have been labelled with specific words. Figure 7 shows the results of such a search: only those sounds labelled "owl" that are most similar to the example owl recording are displayed. A combination search can be used creatively to locate, for example, coyote howls that resemble a siren, or saxophone samples that match an elephant's bellow.
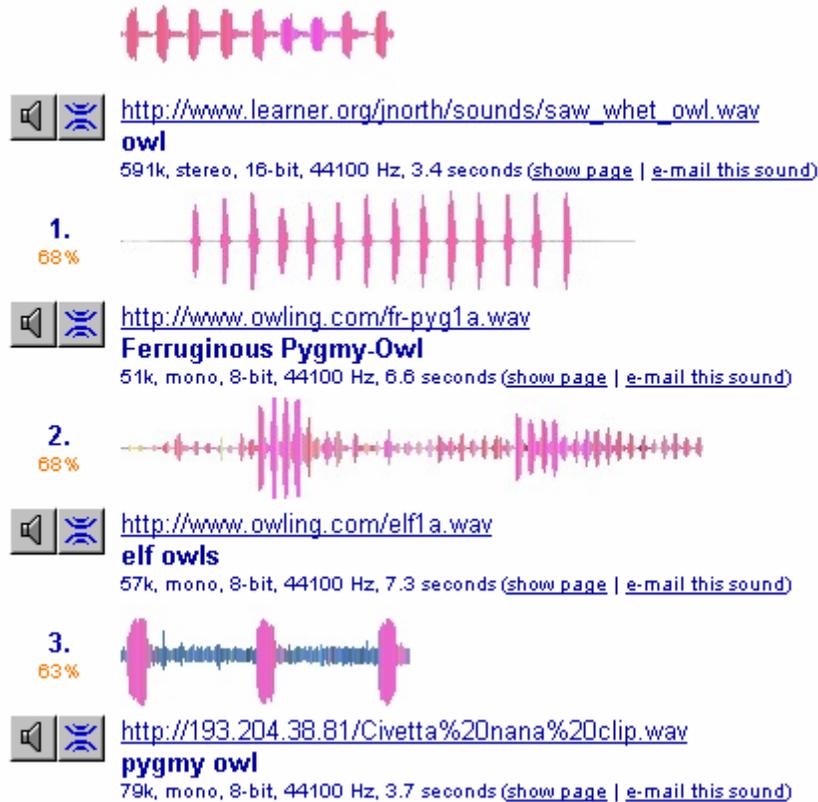
Figure 7 Combination Search Results

4.   CONCLUSION

FindSounds is unique among Web search engines for its focus on sound effects and its use of advanced audio retrieval features, including sounds-like and combination searches, and colored waveform representations of recordings. These features are also available in the FindSounds Palette program [2], which can search audio files stored on local hard drives as well as Web audio files.

5.   REFERENCES

[1] Rice, S. V., "Frequency-Based Coloring of the Waveform Display to Facilitate Audio Editing and Retrieval," presented at the AES 119th Convention, New York, New York, 2005 October 7-10.

[2] Rice, S. V., and S. M. Bailey, "Searching for Sounds: A Demonstration of FindSounds.com and FindSounds Palette," *Proc. of the International Computer Music Conference*, 2004, pp. 215-218.