

Searching for Sounds: A Demonstration of FindSounds.com and FindSounds Palette

Stephen V. Rice* and Stephen M. Bailey†

*Department of Computer and Information Science, The University of Mississippi
rice@cs.olemiss.edu

†Comparisonics Corporation
sbailey@comparisonics.com

Abstract

FindSounds.com is the first Web search engine for sound effects. In addition to keyword-based retrieval of audio files, FindSounds.com provides a “sounds-like search” for content-based retrieval. FindSounds Palette is a unique software program enabling local and remote audio files to be searched by content and at multiple speeds.

1 Introduction

Audio recordings can be categorized into three groups: (1) speech recordings, (2) song recordings, and (3) everything else. For the first group, speech recognition can be applied to convert speech audio to text and this text can be searched to locate speech recordings of interest. For the second group, the burgeoning field of music information retrieval attempts to search collections of song recordings by identifying melodies and musical genres. The third group, for lack of a better name, we shall refer to as “sound effects.” This group includes not only the sounds of explosions and sirens, but also non-speech utterances of the human voice (e.g., a grunt or scream) and musical instrument samples (e.g., a piano note or drum beat).

The ability to search sound effects has traditionally been limited to simple keyword searches of their text descriptions. Human catalogers attempt to describe sound effects in words. Onomatopoeia, the formation of words to imitate sounds (e.g., buzz, thud, pop), has been raised to an art form by catalogers in desperate attempts to describe sounds. For example, in one commercial sound effects library, sounds have been labelled “pingy wobbles,” “heavy zonk,” and “wiggle bowang.”

The cataloger more often resorts to labelling a sound with a description of the source of the sound, if the source is known. A phrase like “elephant trumpeting” or “squealing car brakes” is helpful. However, sound designers developing film soundtracks, and electronic musicians composing with sound samples, are much more interested in the sound than what created the sound.

The limitations of keyword searching have sparked interest in content-based retrieval of sound effects.

Stephen V. Rice, co-author of this article, developed a “sound-matching” algorithm in 1997 for measuring the similarity of sounds and applied this algorithm to implement a “sounds-like search.” Given any sound as input, similar sounds are located automatically in a collection of digital audio files. Other algorithms for this problem have been independently developed, for example, Wold et al (1996) and Foote (1997).

Rice and co-author Stephen M. Bailey began working in 1999 to build a Web search engine for sound effects that incorporates both keyword searching and Rice’s sounds-like search. The result is FindSounds.com which debuted in August 2000. FindSounds.com is the first Web search engine devoted to sound effects.

In its first 3.5 years of operation, FindSounds.com has been visited 7.8 million times and has displayed 46 million page views. It appeals to the general Internet audience but is especially valuable to sound designers, electronic musicians, filmmakers, videographers, and developers of Web sites, computer games, and multimedia applications. This search engine has been profiled on TechTV and in numerous magazines and newspapers (see for example Ananthaswamy 2001; Butner 2001; Notess 2000; O’Connell 2002; Paumgarten 2003; and Zetter 2002).

In Section 2, we describe FindSounds.com from the developers’ perspective, and in Section 3, we introduce FindSounds Palette, a software program that extends the capabilities of FindSounds.com.

2 FindSounds.com

FindSounds.com is a Web search engine like Google but on a smaller scale and with a focus on sound effects. Users perform a keyword search or sounds-like search and receive a list of “hits,” which are links to audio files on the Web. Clicking on a link downloads and plays an audio file.

2.1 Search Features

Keyword searches are specified by entering any word or phrase in a search box, or the user may click on one of the “keyword links” on the “Sound Types” page. For example,

clicking on the “elephant” link launches a search for elephant sounds and is a convenient shortcut for typing “elephant” in the search box. The Sound Types page offers 500 different keyword links organized by category. See Figure 1.

<u>Category</u>	<u>Examples</u>
Animals	alligator, elephant, squirrel, whale
Birds	falcon, kookaburra, swan, turkey
Household	door, kettle, phone, vacuum cleaner
Insects	bee, cicada, cricket, mosquito
Mayhem	broken glass, explosion, gun, sword
Miscellaneous	beep, pop, scrape, squeak
Musical	
Instruments	bass drum, cello, flute, gong, trumpet
Nature	fire, rain, thunder, wind
Noisemakers	alarm, bell, horn, siren
Office	elevator, fax, file cabinet, stapler
People	applause, cough, footsteps, scream
Sports &	
Recreation	basketball, bowling, cards, tennis
Tools	axe, chain saw, jackhammer, torch
Vehicles	helicopter, motorcycle, race car, truck

Figure 1. Categories of sound effects on the Sound Types page at FindSounds.com.



This icon appears beside each hit returned by FindSounds.com. Clicking on it launches a sounds-like search which retrieves up to 200 audio files that sound most like the hit, with the best matches appearing first as determined by the sound-matching algorithm. For each match, a similarity score (e.g., 87%) indicates the degree of similarity between the match and the original sound, and the matches are listed in order of decreasing score.

The matches are determined by the algorithm based on characteristics of their sounds without regard to their source or labelling. Consequently, a search for sounds similar to a tiger’s roar may match the sound of a truck engine. Such a match is of interest to sound designers yet would never be discovered through keyword searches. FindSounds.com makes it possible to discover audio files based on how they sound rather than how they are labelled.

Sometimes the user’s needs are best served by combining a sounds-like search with a keyword search. FindSounds.com enables users to search for audio files that sound like a particular audio file *and* are labelled with one or more keywords. For example, the user can retrieve sounds that are similar to a tiger’s roar and are labelled “tiger” or “lion.”

The sound-matching algorithm emulates the human perception of sound similarity. It analyzes digital audio data, extracts perceptual features, and compares them to estimate the similarity of sounds as perceived by humans. Computers lack ears and human intelligence, so it is a challenge to develop an algorithm that “hears” sounds like

humans. Ultimately, humans are the judge of its accuracy. Users will consider some matches to be better than others and would likely rank the matches in a different order. Nonetheless, the algorithm offers the unique and valuable service of finding similar sounds among thousands of sound effects on the Web.

For all searches (keyword, sounds-like, and combined), users can specify the desired file formats (Wave, AU, and/or AIFF), number of channels (mono and/or stereo), minimum resolution (8-bit or 16-bit), minimum sample rate, and maximum file size of the retrieved audio files.

For each hit, the audio file URL is displayed along with several other file attributes including file size, number of channels, resolution, sample rate, and duration. A short text description of the sound may appear in bold lettering. A link may be provided to a Web page referring to the audio file, and another link makes it easy to e-mail the file’s URL.

Notably, a “Comparisomics waveform display” is shown for each hit. This is an audio waveform display that has been colored to represent the frequency content of the audio data. It is a “thumbnail” image providing visual information about the sound of an audio file. Often simply by inspecting this image, a user can decide whether to download and play the file. This display has proven to be immensely helpful. For more information about this display, see Rice and Patten (2001) and Rice and Latartara (2004).

2.2 Index Creation and Characteristics

FindSounds.com is focused on sound effects of short duration (≤ 10 seconds) stored in Wave, AU, or AIFF format. These are the most commonly-used file formats for sound effects on Windows, Unix, and Macintosh systems, respectively.

FindSounds.com employs a “spider” program to locate audio files on the Web. Each file is downloaded and analyzed by the program to determine if it satisfies the following criteria:

- valid Wave, AU, or AIFF format
- audio data is 8-bit or 16-bit uncompressed, or one of the supported compression types (e.g., ADPCM, A-Law, μ -law)
- mono or stereo
- sample rate between 8,000 and 96,000 Hz
- duration between 0.05 and 10 seconds
- file size of two megabytes or less
- DC offset less than 6.25% (to eliminate poor-quality recordings with excessive DC offsets)
- maximum amplitude greater than 25% (to eliminate poor-quality recordings in which the sound level is too low)

This process automatically eliminates about 90% of the located files. The remaining files proceed to the “auditioning” phase in which humans listen to each file. If a file contains at least one spoken word or a sequence of at least three different notes or chords, then it is considered to

be a speech or song recording and is rejected. Any recording deemed obscene is also rejected. Otherwise, the recording is accepted by the auditioner.

This process is streamlined so that an auditioner may listen to files one after another at a rate of about 1,000 files per hour. Speech recordings account for the vast majority of these and can be rejected upon hearing the first spoken word without listening to the entire file. About 15% of the auditioned files are accepted.

Accepted files advance to the “labelling” phase in which a human cataloger listens to each file and enters a text description of the sound. The cataloger has access to the name of the file, and text from a Web page that links to the file, to assist him in describing the sound. If the sound defies description (as so many do), then it is not labelled. The text descriptions that appear in bold lettering beside a hit are labels entered by the cataloger. These labels are used to answer keyword queries. On average, 58% of the sounds receive labels; the remaining 42% are unlabelled but can be retrieved by a sounds-like search.

Automatic duplicate detection is an essential part of the indexing process. As many as 367 identical copies of a recording have been located on the Web. URLs of copies are saved in a database so that if one copy becomes inaccessible (i.e., the file goes offline), the index can be updated to refer to another copy. Users receive the URL of only one copy in a list of hits so they are not bothered by multiple hits for identical files.

Over its lifetime, the FindSounds spider has located about 10 million audio files on the Web, and about 90% of these were rejected automatically. The remaining one million files, after duplicates are detected, represent about 600,000 different audio recordings. These 600,000 recordings have been auditioned by human listeners who accepted about 100,000 of them for inclusion in the FindSounds index. However, since files on the Web become inaccessible over time, the current number of audio recordings in the FindSounds index is about 50,000.

The average duration of files in the index is 2.8 seconds and the average file size is 105 kilobytes. Wave format dominates (86%) compared with AU (7%) and AIFF (7%). Stereo sound effects are relatively uncommon (13%) compared with mono (87%). Sample rates are distributed as follows:

- 38% less than 12,000 Hz
- 39% between 12,000 and 44,000 Hz
- 23% greater than 44,000 Hz

2.3 Comparison with Other Web Search Engines

Although there are millions of Web sites, there are only about a dozen search engines for finding text, image, audio, or video files on the Web. See Figure 2. Some are famous, like Google and AltaVista, while others are less well known, like FindSounds, and for image retrieval, Picsearch and

Ditto. There are currently only four search engines for locating audio files on the Web: FindSounds, Singingfish, AltaVista, and AlltheWeb.

<u>Name</u>	<u>Locates</u>
AlltheWeb	text, image, audio, and video files
AltaVista	text, image, audio, and video files
Ditto	image files
FindSounds	audio files (sound effects)
Gigablast	text files
Google	text and image files
Picsearch	image files
Singingfish	streaming audio and video
Teoma	text files
Thunderstone	text files
WiseNut	text files
Yahoo!	text and image files

Figure 2. Web search engines.

Singingfish is focused on streaming audio and is therefore not a source for finding individual sound effects. AltaVista and AlltheWeb apparently index all of the audio files they can find on the Web. This inclusiveness is in sharp contrast to the careful selection of sounds for the FindSounds index. Because sound effects are relatively rare on the Web compared with speech and song recordings, it can be difficult for users to locate sound effects in a list of hits obtained from an indiscriminant indexing of audio files. For example, a search for “elephant” may return as many recordings of Henry Mancini’s *Baby Elephant Walk* as elephant sounds.

While it is convenient to build an index in a fully-automated manner like AlltheWeb, human involvement produces a FindSounds index of sound effects and only sound effects, with better text labels. Users receive only sound effects in response to queries and enjoy the categorization of sound effects provided by the Sound Types page. Conceptually, the semi-automated process of creating the FindSounds index is somewhere between the fully-automated construction of the AlltheWeb index and the fully-manual development of the LookSmart directory of Web sites.

3 FindSounds Palette

FindSounds Palette is a state-of-the-art audio retrieval system. It is an audio player, recorder, editor, database, search engine, and Web browser, all in one software program. It is intended for people who collect or need access to sounds, or wish to create and edit sounds. Version 1.0 debuted in September 2002; Version 2.0 was released in January 2003.

With this program, users can catalog and search audio files on their local hard drives and local area network. A database called “MyPalette” contains information about local audio files. The user may enter the following metadata

into MyPalette for each local audio file: description, source, copyright, notes, genre, key, and tempo. In addition, each file may be placed in a class (Effect, Instrument, or Other) and in a system-defined or user-defined category and sub-category. The main window of the program displays a hierarchical view of MyPalette files organized by class, category, and sub-category.

The collection of sounds at FindSounds.com is accessible using FindSounds Palette and is called "WebPalette." With one query, a user can search MyPalette and WebPalette to find local and remote files satisfying search criteria. Up to 200 MyPalette hits are returned in one list, and up to 200 WebPalette hits are retrieved in another. Hits may be displayed in tabular form and sorted by clicking on column headings, or may appear in a "stacked" view like the hits at FindSounds.com. A Comparisonics waveform display appears beside each hit along with icons for playing the file, opening the file in the audio editor, and launching a sounds-like search. Once opened in the audio editor, a WebPalette file can be saved locally to MyPalette.

Users can specify keyword, sounds-like, and combined searches. Any search may place restrictions on file format, file size, number of channels, resolution, sample rate, duration, key, and tempo. Keyword searches can apply to any combination of text fields: file name, description, source, copyright, notes, genre, category, and sub-category. A sounds-like search can specify the desired range of similarity scores.

Hollywood sound designers know well that useful and interesting sounds can be obtained by changing the playback speed of an audio recording. For example, in his book on sound effects, Mott (1990) describes how a single recording of a waterfall, when played at different speeds, has been used convincingly to create the sounds of printing presses and atomic bomb explosions.

In a sounds-like search, FindSounds Palette can search the sounds produced by playing audio files at various speeds. Each file in MyPalette can be indexed at its normal speed and 24 other speeds: the normal speed modified by plus or minus one to 12 semitones. This has the effect of multiplying the size of the local audio collection, but without increasing hard disk utilization because each audio file is stored only once, at its normal speed. A collection of 10,000 local audio files thereby becomes a searchable database of 250,000 sounds. Likewise, each WebPalette file is indexed at more than 40 speeds. The 50,000 sounds in the FindSounds index become a searchable collection of 2,000,000 sounds!

The audio editor enables a user to play, record, and edit an audio file while viewing its Comparisonics waveform display. Editing operations include cut, copy, paste, mix, delete, fade, adjust volume, undo, and redo. The user may change the playback speed of a file, and appropriately, the waveform display is automatically repainted with new colors to reflect the change in sound.

The user may select any sound in an audio file by highlighting it in the colored waveform display. Clicking on the sounds-like search icon retrieves sounds in MyPalette and WebPalette that are similar to the selected sound. Using his voice or props, a user can mimic a desired sound into a microphone and find similar sounds available locally and on the Web.

FindSounds Palette is available for computers running Microsoft Windows. A free trial can be downloaded from: <http://www.FindSounds.com/Palette.html>

4 Conclusion

FindSounds.com offers unprecedented access to sound effects on the Web. Keyword searching supplemented with content-based retrieval, hits illustrated by colored waveform displays, and careful semi-automated index construction, create a powerful and enjoyable Web search engine. FindSounds Palette extends the search capabilities of FindSounds.com to local audio files, enables sounds-like searching of multiple speeds, and integrates a unique colored waveform editor.

References

- Ananthaswamy, A. (2001). "You Hum and I'll Find It." *New Scientist*, March 17, 34–37.
- Butner, R. (2001). "Incredibly Useful Sites." *Yahoo! Internet Life*, January, 92–94.
- Foote, J. T. (1997). "Content-Based Retrieval of Music and Audio." In *Multimedia Storage and Archiving Systems II, Proceedings of SPIE 3229*, 138–147.
- Mott, R. L. (1990). *Sound Effects: Radio, TV, and Film*. Boston: Focal Press.
- Notess, G. R. (2000). "Searching Beyond Text: Multimedia Search Tools." *Online*, November/December, 63–65.
- O'Connell, P. L. (2002). "Armchair Movie Criticism and Sound Searches." *New York Times*, May 9.
- Paumgarten, N. (2003). "A Rare and Different Tune." *The New Yorker*, September 15, 36–38.
- Rice, S. V., and J. Latartara. (2004). "Frequency-Based Coloring and Navigation of the Audio Waveform Display." Technical report TR-2004-60, Dept. of Computer and Information Science, Univ. of Mississippi.
- Rice, S. V., and M. D. Patten. (2001). "Waveform Display Utilizing Frequency-Based Coloring and Navigation." U.S. patent 6,184,898.
- Wold, E., T. Blum, D. Keislar, and J. Wheaton. (1996). "Content-Based Classification, Search, and Retrieval of Audio." *IEEE MultiMedia* 3(3), 27–36.
- Zetter, K. (2002). "Best of Today's Web: Greatest Hits and Hidden Gems." *PC World*, August, 69–78.