New software tool provides unprecedented searches of sound, from musical riffs to gunshots

November 9 2011

The algorithms driving Imagine research's MediaMinedTM software differentiate between instruments, voices and other sounds without needing keywords and indexes sound files to allow for sound-similarity searches. In the image above, the sound file for Yoda's quote "You must feel the force around you ..." with background music and noise (left screen) looks quite different from the sound file for an acoustic guitar (center screen) and that for an electric guitar (right screen). However, the true impact of the software is its ability to find clips that relate to each of those sound files based on a user's need. For example, a user can search for hard-rock tracks with extensive electric guitar solos without weeding through acoustic guitar or even jazz guitar tracks -- all without ever entering a keyword. Credit: Imagine Research, Inc

Audio engineers have developed a novel artificial intelligence system for understanding and indexing sound, a unique tool for both finding and matching previously un-labeled audio files.

Having concluded beta testing with one of the world's largest Hollywood [sound](#) studios and leading media streaming and hosting services, Imagine Research of San Francisco, Calif., is now releasing MediaMined for applications ranging from music composition to healthcare.

The company developed the tool with support from the National Science Foundation's Small Business Innovation Research program (IIP-0912981 and IIP-1206435).

"MediaMinedTM adds a set of ears to cloud computing," says Imagine Research's founder and CEO Jay LeBoeuf. "It allows computers to index, understand and search sound--as a result, we have made millions of media files searchable."

For recording artists and others in music production, MediaMinedTM enables quick scanning for a large set of tracks and recordings, automatically labeling the inputs.

"It acts as a virtual studio engineer," says LeBoeuf, as it chooses tracks with features that best match qualities the user defines as ideal. "If your software detects male vocals," LeBoeuf adds, "then it would also respond by labeling the tracks and acting as intelligent studio assistant--this allows musicians and audio engineers to concentrate on the creative process rather than the mundane steps of configuring hardware and software."

For special effects studios, MediaMinedTM offers a new approach to sound searches. "Let's say you are working on a movie, and the director needs some explosions," says LeBoeuf. "The state of the art for searching for sounds in multi-terabyte audio collections is to search on the text--usually the filename--of the sounds. So, the sound editor could find 'explosion'--but would never find tracks that were labelled 'big bang', 'huge blast', 'detonation', 'nuclear blast', 'bomb', etc.

MediaMinedTM is capable of grouping those sounds together--you would give us an example of what you are looking for (the sound of an explosion) and we are able to return things that sound like an explosion--regardless of their underlying metadata, name or text content."

The technology uses three tiers of analysis to process audio files. First, the software detects the properties of the complex sound wave represented by an audio file's data. The raw data contains a wide range of information, from simple amplitude values to the specific frequencies that form the sound. The data also reveals more musical information, such as the timing, timbre and spatial positioning of sound events.

In the second stage of processing, the software applies statistical techniques to estimate how the characteristics of the sound file might relate to other sound files. For example, the software looks at the patterns represented by the sound wave in relation to data from sound files already in the MediaMinedTM database, the degree to how that sound wave may differ from others, and specific characteristics such as component pitches, peak volume levels, tempo and rhythm.

Because the software's sound database continues to grow--it currently contains over two million files totaling ten terabytes--the characterization ability of the software continues to improve as the product attracts more users and analyzes additional files.

In the final stage of processing, a number of machine learning processes and other analysis tools assign various labels to the sound wave file and output a user-friendly breakdown. The output delineates the actual contents of the file, such as male speech, applause or rock music. The third stage of processing also highlights which parts of a sound file are representing which components, such as when a snare drum hits or when a vocalist starts singing lyrics.

"MediaMinedTM listens to audio files that are uploaded to our servers, and we generate an XML output with the low-level perceptual content, a universal sound signature and a high-level description of the audio in the file," says LeBoeuf. "When software applications understand what they are listening to, they can do a better job processing audio and help users discover new content."

One of the key innovations of the new technology is the ability to perform sound-similarity searches. Now, when a musician wants a track with a matching feel to mix into a song, or an audio engineer wants a slightly different sound effect to work into a film, the process can be as simple as uploading an example file and browsing the detected matches.

"There are many tools to analyze and index sound, but the novel, machine-learning approach of MediaMinedTM was one reason we felt the technology could prove important," says Errol Arkilic, the NSF program director who helped oversee the Imagine Research grants. "The software enables users to go beyond finding unique objects, allowing similarity searches--free of the burden of keywords--that generate previously hidden connections and potentially present entirely new applications."

While new applications continue to emerge, the developers believe MediaMinedTM may aid not only with new audio creation in the music and film industries, but also help with other, more complex tasks. For example, the technology could be used to enable mobile devices to detect their acoustic surrounding and enable new means of interaction. Or, physicians could use the system to collect data on such sounds as coughing, sneezing or snoring and not only characterize the qualities of such sounds, but also measure duration, frequency and intensity. Such information could potentially aid disease diagnosis and guide treatment.

"Teaching computers how to listen is an incredibly complex problem,

and we've only scratched the surface," says LeBoeuf. "We will be working with our launch partners to enable intelligent audio-aware software, apps and searchable media collections."