Open Fingerprint™ Architecture Whitepaper
Version 1.0, March 12th, 2006

Table of Contents

Overview

This whitepaper documents the design goals, methodology, infrastructure, and validation of the Open Fingerprint™ acoustic fingerprint. The aim of this document is explanatory, for both technical and non-technical audiences. Statistics, tuning of algorithms and other aspects of the fingerprint in use may evolve over time. This document will be updated as the technical specifications of the print evolve.

What is an acoustic fingerprint? "…a unique code generated from an audio waveform." *(*Wikipedia 1/14/2006). In simple language, the purpose of an acoustic fingerprint in digital music is to consistently and rigorously identify the sounds in an audio file regardless of variations in the digital-level details.

The Open Fingerprint™ acoustic fingerprint uses a combination of client-side processing of acoustic signals, and server-side resolution against an extremely large dataset (>14 Million as of January 2006) existing fingerprints to achieve rigorous, consistent identification of identical master recordings.

The Open Fingerprint algorithms, codes, and supporting technologies were developed by Predixis Corporation of Monrovia, California during 2000-2005.

The designated Web site for community support of the open source code is www.musicdns.org.

Design Goals

The Open Fingerprint™ algorithms and technology architecture are not general-purpose. This section outlines the design goals they aim to achieve to clarify the domains where they should—and should not—be applied.

Digital music

First and foremost, Open Fingerprint was designed for use with digital music

Open Fingerprint is not designed for voice recognition or non-music sound samples.

More specifically, Open Fingerprint is designed and tempered for accurate recognition of recorded music—that is, the kind of musical 'content' that people listen to by choice and for emotional engagement.

Open Fingerprint is not designed to work with non-recorded kinds of digital music—such as midi files, synth patches, samples, and so on.

Open Fingerprint is song-centric; it aims to identify individual tracks of music.

Open Fingerprint is not aimed at picking individual songs out of mash-ups or run-on recordings.

Because Open Fingerprint was designed to identify the same piece of music, consistently, anywhere in the world, it's properly called a "tight" fingerprint.

Digital Independence

Open Fingerprint is designed to bridge the seeming contradictions of dealing with digitized music while overlooking the digital details. As a result of careful R&D, the Open Fingerprint manages the following:

- Format Independent. Digital music shows up in many formats—not only the ubiquitous MP3, but WMA, AAC, Ogg, Flac, Lame, AAC+. Open Fingerprint has a proven track record with every format in common use.
- Bit-rate Independent. Open Fingerprint has been validated with the lowest bit rates humans can tolerate (as low as 64K), and the highest-bit-rate digital recordings available.
- Loss Independent. Open Fingerprint operates equally well on "lossy" formats such as MP3 and WMA, and "lossless" formats like FLAC or WAV.
- DRM-independent. Open Fingerprint is blind to DRM issues. If the device running the fingerprint algorithms can play the song, Open Fingerprint will identify it.

Metadata Independent

The Open Fingerprint does not rely on the non-acoustic information in (or associated with) a track to identify the track.

Whole-Song 1-Track Centric

Open Fingerprint was designed for song-by-song use. As personal computers, in general, handle music on a one-file-for-one-track basis, current implementations of the Open Fingerprint code are designed to process files on the assumption that they are single songs. The same algorithms could be used to analyze streamed music, or deliberately-chunked 'whole song' pieces from larger files, but the services and code are not tuned that way at present.

In other words: Open Fingerprint is designed to identify songs, and currently implemented to handle files.

Identification Integrity, Not Rapid Recognition

The goals of quickly identifying a song from a partial fragment, and rigorously identifying a song, have inherent opposition.

There are acoustic fingerprints on the market that were designed for rapid recognition—all they require is a fragment of a track to identify it out of a set of known tracks. (The best-known example of use is probably 'hold up your phone and we'll name the track.')

While this is an interesting application, the tempering of the print and of the data-set required to accomplish this run counter to goals of rigor. These technologies work best with a very limited song-set (e.g. popular music from the radio); logically, small fragments are harder to accurately recognize from large song-sets.

In addition, partial-fragment approaches are more easily spoofed or evaded. All a 'hacker' need do is tack a fragment from a public-domain track onto the front of protected content to evade detection.

By contrast, the Open Fingerprint design goal is rigorous identification, with integrity and consistency across the common variables of use noted elsewhere.

Tempered

Artifacts of the early stages of networked digital music affect the design requirements of the Open Fingerprint. As compression enabled smaller files and conserved bandwidth, early conversions from analog and CD digital were frequently very poor in quality. Naturally, original recording constraints—mikes, tapes, masters, etc.—had their own effect on audio quality.

The need to see past these artifacts and consistently recognize tracks is another key design goal for the Open Fingerprint algorithms and technical implementation.

False Positives / False Negatives

A false positive, also called a Type I error, exists when a test incorrectly reports that it has found a result where none really exists. (Wikipedia, January 2006). In acoustic fingerprinting, identifying a track as "Born To Run", when it is in fact "Knuckle Sandwich" , would constitute a false positive.

A false negative, also called a Type II error or miss, exists when a test incorrectly reports that a result was not detected, when it was really present. (ibid). In acoustic fingerprinting, not identifying a track as "Born To Run", when it is in fact "Born To Run", would constitute a false negative.

The Open Fingerprint algorithms, and the MusicDNS data-set, are tightly tuned to avoid both false positives and false negatives.

All The Music In The World

The Open Fingerprint is designed to work successfully with a very large and unknown-in-advance data set—that of all the possible music in the world. There's no predicting what songs will show up in the world—and what derivations, cover bands, mash-ups and other variations that might be included. Rather than being tempered for the special purpose of a small, predictable set of popular tracks, this print and architecture is designed to discern fine details, and to be tuned and evolved as the data-set grows without changing any distributed code.

MusicIP, 605 E. Huntington Drive, Suite 201, Monrovia, CA 91016
Phone 626-359-9702 • Fax 626-359-9827 • http://www.predixis.com

3. Methodology

Generating the Open Fingerprint is a 3-step process. The final outcome is a set of numbers (a 516-byte array) handled by the MusicDNS server to return an song ID resolved against the 'world' song-set.

These steps assume two things, both important to note in understanding performance and identification rigor.

1) digital files have to be decoded for fingerprinting to take place.

2) the Open Fingerprint is standardized on 2 minutes, or the full song length if <2 minutes.

Although the Open Fingerprint is extremely lightweight (<2000 lines of code) and fast (seconds per track), decoding a track imposes unavoidable overhead. It's conceivable that fingerprinting could be done in conjunction with some other decode process (e.g. playing the file), but current implementations handle decode directly. Decoding is unavoidable, by definition.

The 2-minute file length is considerably longer than the time-sample on which most commercial fingerprints were calibrated, by design. The length was designed to cover the majority of a track. Across our sample of 14,000,000, tracks average 3.2 minutes in length. Although a 'hacker' could in theory tack 2 minutes of audio information in front of a song, the likelihood of listeners tolerating the noise is very low.

Normalization

After decoding, the first functional stage of the fingerprint is a set of preprocessing steps which normalize the signal to a standard set of properties.

Frequency Extraction

After normalization, the Open Fingerprint calls an FFT (Fast Fourier Transform) from an external library. *"In signal processing and related fields, the Fourier transform is typically thought of as decomposing a signal into its component frequencies and their amplitudes."* (Wikipedia, January 2006) The FFT libraries tested and validated with the Open Fingerprint include:

- the FFTW (Fastest Fourier Transform in the West), an academic-release library release, available for free use under GPL or licensable separately for commercial use. (http://www.fftw.org/)

- the Intel Math Kernel Library (http://www.intel.com/cd/software/products/asmo-na/eng/perflib/mkl/index.htm), a high-performance, low-cost FFT library.

Use of Open Fingerprint with the MusicDNS service requires use of one of these libraries. Applications using other FFT libraries must pass the test harness jointly

managed by MusicBrainz and Predixis, and obtain a modification to their terms of service prior to using MusicDNS. (Other libraries may return incorrect results.)

The Open Fingerprint operates on the amplitude data returned by the FFT. It examines a series of spectra, each representing a small frame (185 milliseconds) of the audio. Each frame in turn consists of a number of frequency "bins" that hold the value for the amplitude of a particular frequency band. Open Fingerprint treats these as a matrix, with rows corresponding to time and columns corresponding to frequency.

Singular Value Decomposition

Given a matrix such as this, there exists a mathematical equation called the Singular Value Decomposition.

"For an m-by-n matrix A with m >= n, the singular value decomposition is an m-by-n orthogonal matrix U, an n-by-n diagonal matrix S, and an n-by-n orthogonal matrix V so that A = U*S*V'.

The singular values, sigma[k] = S[k][k], are ordered so that sigma[0] >= sigma[1] >= ... >= sigma[n-1]." (from the JAMA documentation)

The result of the SVD operation on the amplitude data from the audio frames is, effectively, a much smaller matrix that nonetheless maximizes the information transferred from the large matrix.

The resulting set of vectors—the "core print" of the Open Fingerprint—comprises 512 bytes, a relatively small amount of information for transmission and further use.

Pitch Print

The Open Fingerprint incorporates a final operation which captures a different perspective of the acoustic information in a file. From the spectrum matrix described previously, the Open Fingerprint algorithms identify "peak-trajectories", which are prominent frequencies that have continuities from frame to frame. These are measured cumulatively to build up a ranking of the strongest pitches encountered. The four most prominent pitches are returned as byte values.

These pitch print values are used to narrow the lookup set for fingerprint matching and ID return.

4. Implementation

The contextual data-set for the Open Fingerprint is large—estimates of the number of master recordings in existence run to 25 million and up. It is also unknown. Consequently, to meet the design goal "all the music in the world", the implementation of the Open Fingerprint includes a final distinction and 'tuning' operation, conducted against the cumulative set of prints in existence.

In simple terms, this means that the core print, plus the pitch print, are matched to existing prints, at a central server with access to the entire set. Furthermore, this server is capable of being tuned to ever-tighter tolerances as the set of music grows. (There is, to be clear, no mathematical danger of 'running out of space.')

Fingerprint Transfer

Transporting the SVD matrix and the pitch verification bytes to the server is straightforward. The 516-byte matrix (512 x SVD, 4 x pitch verification) is BASE-64 encoded. *"…base64 is a binary to text encoding scheme whereby an arbitrary sequence of bytes is converted to a sequence of printable ASCII characters."* (Wikipedia, January 2006) The resulting 565-byte string is submitted to the MusicDNS server farm via HTTP.

Note: as of this whitepaper, the credential scheme for access to the MusicDNS service was still under design.

Fingerprint Resolution

The MusicDNS server farm, built on top of the proven Predixis fingerprint servers, provides extremely rapid resolution and lookup of single fingerprints from the current set globally-known prints. The servers use a distributed-query N-tier architecture, similar in design to large search engines (e.g. Google.) Core print lookup information is held in RAM memory, enabling the fastest possible return times. Current benchmarks return approximately 100 fingerprint lookup returns per second per server.

Resolving the fingerprint and returning a reliable song ID is a complex but fast mathematical operation, involving comparison of the SVD vectors against existing SVD vectors, using server-tuned tolerances.

XML Document

The final step in fingerprinting is the return of a reliable, useful (small enough to handle quickly) canonical ID, along with metadata about the track. As of this whitepaper, the final design for the XML document standard for returning track ID and metadata is under discussion. At a minimum the fields will include

- a track ID, in the form of a 32-byte GUID-like string
- basic public-domain metadata, including Artist, Album and Track name(s)

At the risk of stating the obvious, the track IDs are the reliable element resulting from the fingerprint process. Metadata can and will change.

Subsequent information on the metadata schema will be posted at www.musicdns.org at release of the service.

MusicIP, 605 E. Huntington Drive, Suite 201, Monrovia, CA 91016
Phone 626-359-9702 • Fax 626-359-9827 • http://www.predixis.com