Julia C. Thompson. Evaluating User Perspectives of Audio Fingerprinting Technologies. A Master's Paper for the M.S. in L.S degree. December, 2008. 34 pages. Advisor: Jane Greenberg.

Audio fingerprinting, the process by which an audio sample is automatically identified or categorized based on its unique analog properties, is a technology that has been integrated quite rapidly into commercial and consumer music applications. Although academic literature has covered specific algorithms, technical specifications of prototype applications, and practical implementations of audio fingerprinting software, little attention has been given to public opinions. However, public opinion has been debated extensively in online blogs, forums, newspapers, and e-mail lists. This study addresses this knowledge gap by examining a sample, via content analysis, of 30 web-based resources, revealing public opinions about this topic. The sample was drawn from alternative resources, as they provide a rich information source for understanding public perception and opinion about audio fingerprinting. This paper presents background research and key new findings based on this study. The content analysis identified key concepts and unifying discussion themes. The research approach and analysis was verified by two independent evaluators, confirming consistency in coding. The results revealed significant public interest in topics of audio fingerprinting metadata, and emphasized the following themes as important topics among the public: 1. user interfaces, 2. technical issues, 3. copyright implications, 4. royalties, and 5. user privacy. The paper concludes with a series of recommendations for future research.

Headings:

Audio fingerprinting

User-centered system design

Music information retrieval

Music/Internet resources

Content-based music retrieval

Content-based analysis

EVALUATING USER PERSPECTIVES OF AUDIO FINGERPRINTING TECHNOLOGIES

by
Julia C. Thompson

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

December 2008

Approved by

____

Jane Greenberg

TABLE OF CONTENTS

INTRODUCTION

As digital audio becomes more pervasive in everyday society, users constantly search for ways to improve access, organization, and identification. In recent years, audio fingerprinting technologies have been applied to address these needs and related issues in both consumer and non-consumer markets.

An audio fingerprint (synonymous with “acoustic fingerprint”) is a digital measure of an audio file’s analog properties, and can be used to identify a unique audio sample and quickly locate or categorize the sample in an audio database.[1] There are two main processes to any audio fingerprinting system. When an audio file is presented, the first process computes the fingerprint from the file. The second process uses sophisticated search algorithms to scan a database of previously computed fingerprints for matches. A robust audio fingerprint algorithm takes into account the analog characteristics of audio. If two files sound alike to the human ear, their acoustic fingerprints will be equal, or very similar, even if their binary representations are different. Most audio compression techniques (MP3, OGG Vorbis, etc.) make radical changes to the binary coding of an audio file. However, a robust audio fingerprint will allow a recording to be identified

1 “Audio fingerprinting” differs from “Audio watermarking,” where additional information is imbedded into the original signal. This mark is generally imperceptible to human ears, but is easily distinguished by robust identification technologies. The processes for these identification technologies differ significantly from those used for audio fingerprinting. For the purposes of this study, only audio fingerprinting will be discussed.

after it has gone through such compression, even if the audio quality has been reduced significantly.

Much like human fingerprints, no two audio recordings are alike. For example, take two different recordings of the Beatles' "Lucy in the Sky with Diamonds." One recording is the commercially-produced version from Abbey Road Studios, released on the album: *Sgt. Pepper's Lonely Hearts Club Band*, while the other is a live recording from a concert in Liverpool. When using your favorite digital media player application (iTunes, VLC, etc.), these two recordings can be identically tagged and labeled. However, upon listening to each, it is clear that the recordings are different. When audio fingerprinting software is applied (and assuming that both versions are in the database, and their metadata is correct), the files will be respectively identified as separate recordings of the same piece. (E.g. "Beatles - 'Lucy in the Sky with Diamonds' (Album: *Sgt. Pepper's Lonely Hearts Club Band*)" and "Beatles – 'Lucy in the Sky with Diamonds' (Live Recording: Liverpool, June 6, 1968."))

Practical uses include: Broadcast monitoring, identification of music and ads being played, peer to peer network monitoring, video identification, duplicate song detection in personal libraries, copyright enforcement, data restoration and repair (i.e. restoring metadata), collaborative analysis, song recommendation, and others. Over the past decade, audio fingerprinting techniques have been extensively covered in the computer science and information retrieval literatures. Here, specific algorithms, technical specifications for prototype software, and practical implementations of audio fingerprinting software have been discussed.

While the technical and practical elements have been discussed in literature and refined based on analyses and feedback, the public opinion has scarcely been covered in the literatures. This provokes such questions as: How are users responding to this software as it becomes mainstreamed into their favorite audio software programs? What are their likes/dislikes? What are their suggestions for improvement? Although these questions have not been addressed in academic and research literature, they have been debated extensively in online blogs, forums, and newspapers. These alternative resources thus provide a rich information source for understanding public perception and opinion about audio fingerprinting, and are an excellent starting point for learning about the issues associated with this technology. The research reported in this paper enhances our knowledge of audio fingerprinting by examining these resources (blogs, forums, etc.) and obtaining a more complete view of their development and adoption. This paper reports on a content analysis examining current issues discussed and makes recommendations for future research in the developing area of audio fingerprinting.

BACKGROUND AND LITERATURE REVIEW

System Design and Algorithms

The notion of retrieving music from a database using purely acoustic input has been present since the 1970s (Jayant, 1984). However, it was not until the mid-1990s that implementation became more of a reality. McNab et al. (1996) describes a prototype system that was a significant precursor to subsequent audio fingerprinting programs. The authors experimented with a database containing sheet music for nine thousand, six hundred folksongs. The opening notes of the song were hummed into a microphone, and results were displayed in close-match format. Since then, literally hundreds of prototype fingerprinting algorithms and systems have been postulated. While the evolution and particulars of system design are too broad to document in the scope of this study, adequate synopses can be found in the writings of Cano et al. (2002) and Cano (2007).

However, there is significant consensus on effective design of audio fingerprinting programs. Haitsma and Kalker (2003, pg. 211) state that:

> "The prime objective of multimedia fingerprinting is an efficient mechanism to establish the perceptual equality of two multimedia objects: not by comparing the (typically large) objects themselves, but by comparing the associated fingerprints (small by design)… the fingerprints of a large number of multimedia objects, along with their associated metadata (e.g. name of artist, title and album) are stored in a database. The fingerprints serve as an index to the metadata. The metadata of unidentified multimedia content are then retrieved by computing a fingerprint and using this as a query in the fingerprint/meta-data database. "

A key advantage of a fingerprint database, versus a database of full-size items, lies in the fact that the reduced memory and storage requirements for fingerprints are relatively small. This leads to greater speed and efficiency in searching and accessing information.

Drawing from the writings of Haitsma and Kalker (2003) and Cano et al. (2002), the following parameters have been outlined for proper system design:

1. Accuracy – How often is a song correctly identified? Accuracy refers to the number of correct identifications, missed identifications, and wrong identifications (false positives).

2. Reliability – Tying with accuracy, can the system consistently produce accurate identifications? This is particularly important when software is implemented for copyright enforcement and broadcast monitoring. Incorrect identification can lead to loss in revenue.

3. Robustness – Refers to the ability to accurately identify an item, regardless of the level of compression, distortion, or interference in the transmission channel. Preferably, a severely degraded audio signal can yield an accurate and viable fingerprint.

4. Granularity/Cropping - How many seconds of audio are needed to identify an audio clip – all or part of the song?

5. Search Speed and Scalability – How long does it take to find a fingerprint in a database? What if the database contains millions of songs? How much storage is needed for the fingerprints? For commercial use and other high-volume use, these issues are of particular concern.

6. Security – How susceptible is the system to cracking or tampering?

7. Versatility – Refers to the ability to identify audio regardless of audio format, as well as the ability to use the same fingerprint database for different applications.

As with most concerns of system design, these parameters can have an impact on each other. Haitsma and Kalker (2003, pg. 13) use the example that “if one wants a lower granularity, one needs to extract a larger fingerprint to obtain the same reliability. This is due to the fact that the false positive rate is inversely related to the fingerprint size.” All of these factors should be taken into account when designing a fingerprinting system.

When designing for commercial use, this becomes of prime importance as cost benefit analysis becomes a factor in regards to extraction processes, complexity, and time (Cano, 2007).

Refinements of fingerprint systems are ever-evolving, but the regards and standards for system design remain consistent.

Current Implementations and Applications

As the refinement of audio fingerprinting systems has continued, users have seen widespread implementations in a number of applications, both for commercial and non-commercial uses.

Perhaps the most prevalent use of audio fingerprinting is in broadcast and intellectual property monitoring. In this process, broadcast monitoring programs “listen” to a television, radio, or web station and continually update playlists for purposes of royalty collection, program verification, and advertisement verification. The American Society of Composers and Publishers’ MediaGuide is a prime example of this type of software (MediaGuide, Inc., 2008). This implementation presents a useful alternative to employing people around the clock who monitor station programming. Other companies include Nielsen Broadcast Data Systems (Nielsen BDS, 2008), which is one of the world’s leaders in entertainment tracking, and Audible Magic, which is currently used by YouTube and MySpace for monitoring (Audible Magic Corporation, 2008).

Fingerprinting is also present in consumer applications. One example is the Shazam application (Baluja and Covell, 2008; Shazam Ltd., 2008), designed for cell phones and other mobile devices. One merely has to hear a song playing over the radio, loudspeaker, or other sound source. The phone is then held up to the source, and the sound is processed through the phone's microphone. Within moments, the song is processed through Shazam's database and identified. When used on Apple's iPhone, one has the ability to then purchase the song through Apple's proprietary iTunes program and also share the song with friends and contacts.

Einhorn and Rosenblatt (2005) discussed audio fingerprinting's role in the monitoring of peer-to-peer networks. Drawing from the infamous example of Napster, which was introduced in June 1999, users who downloaded the Napster client shared and downloaded a large collection of music for free. In early 2001, after several court cases initiated by the recording industry, Napster users were forbidden to download copyrighted songs. In March 2001, Napster installed a filter to block the filenames of corresponding copyrighted songs. Users caught on to this quickly and intentionally misspelled filenames to circumnavigate the filter. In May 2001, Napster introduced an audio fingerprinting system by Relatable, which aimed at filtering out copyrighted material despite misspellings in tags and metadata. Shortly thereafter, in June 2001, Napster was closed. (Cano, 2007)

This implementation shows little signs of slowing down. Shrethsta and Kalker (2004) described a fingerprinting system which could be distributed over several computers, so as to make the process much more efficient. One example of this was the popular social networking website MySpace's 2006 implementation of Gracenote's MusicID, which sought to prevent unauthorized copyrighted music from being posted to MySpace user's pages. (Hefflinger, 2006)

Haitsma and Kalker (2003) state that, “from a consumer standpoint, audio filtering could be viewed as a negative technology… but there are also a number of potential benefits to the consumer. First, filtering can organize music song titles in a consistent way, by using the reliable metadata of the fingerprint database. Secondly, fingerprinting can guarantee that what is downloaded is actually what it says it is. (p. 13)”

On the consumer level, many users’ personal libraries contain thousands of songs. The music is generally stored in compressed format on a hard drive or personal audio device (i.e. iPod or other MP3 player). When these songs are obtained from different sources, such as ripping from a compact disc or downloading from a file-sharing network, the metadata is often inconsistent, incomplete, missing, and sometimes incorrect. Assuming that the fingerprint database contains correct metadata, audio fingerprinting can correct the library’s song metadata, thereby facilitating easier organization and access. Two current examples of this are MusicBrainz (MusicBrainz Foundation, 2008) and Gracenote MusicID (Gracenote Inc., 2008). With MusicBrainz, one downloads a program that extracts fingerprints and submits them to a central server, where the metadata for the associated tracks is returned. Gracenote’s database works in much the same way, but downloads also include album art.

One consumer tool for which audio fingerprinting can be used is duplicate song detection. Burges et al. (2005) discuss a system design in which duplicate songs can be identified in a set even if they differ in compression quality or duration. One current implementation is Barcelona Music and Technologies’ Vericast program, which offers this feature. (Barcelona Music and Technologies, 2008)

Finally, another current consumer implementation of audio fingerprinting software is intended for song recommendation. Mufin, Pandora, and MusicIP Mixer currently offer music recommendation systems (Mufin GmBH, 2008; Pandora Media, Inc., 2008; MusicIP, 2008). MusicIP Mixer, for example, uses audio fingerprinting to generate playlists based on audio characteristics of a song or songs selected. These characteristics include “genre” and “style,” as well as various other descriptions about the audio such as “smooth,” “upbeat,” “groovy,” etc. The user also has the option to specify how much emphasis the program places on user-generated tags, rather than on the audio fingerprint itself. While song recommendation is still very much in its infancy, it could become a viable implementation of audio fingerprinting technology in the near future.

User Studies

While user studies have been mostly neglected in regards to audio fingerprinting, they have been discussed in the music information retrieval literature. Downie and Cunningham (2002) analyzed a set of music-related information requests posted to a Usenet newsgroup dedicated to discussion about “Old Time” music. The postings were categorized by the occupation of each writer, of which there was much variety: librarians, musicologists, engineers, lawyers, etc. Postings were coded by the information needs, the type of music information requested, the intended uses for the information, and additional social, environmental, and contextual elements present in the postings. One benefit of this unobtrusive study was that users posted natural language requests, expressing their information needs in their own words and not constrained by search syntax. These requests provided additional insight into context and motivations behind music information retrieval needs and were thought to be a successful exercise towards designing music information retrieval systems that are user-oriented.

Bainbridge, Cunningham, and Downie (2003) examined how people “describe what they want” when expressing their musical needs. Building upon Downie and Cunningham’s (2002) previous study, the latter study examined a larger number of music queries through (the now-defunct) Google Answers service. The investigators examined the language that people used to describe their information requests. Their study set found that over 80% of users used some form of bibliographic metadata to describe their request (performer, title, date, label, etc.), while other requests were much more nebulous (i.e. “The song I’m looking for sounds kind of twangy.” Or “I think the lyrics go something like this: Yada yada yoooo.”)

As outlined by Cunningham, Reeves, and Britland (2003), as of the early 2000’s, only a handful of user studies had been conducted in the music information retrieval/music digital library domain. Thus, many existing music information retrieval systems had been designed and evaluated largely based on anecdotal evidence of user needs, intuitive feelings for user information-seeking behavior, and deductive assumptions of typical usage. As such, Downie and Ha Lee (2004) initiated surveys to “acquire information to help eradicate false assumptions in designing music information retrieval systems.” As of the date of this article’s publication, user studies focusing upon real-life music information needs, uses, and seeking behaviors were still very scarce in the music information retrieval and digital library fields. Downie and Ha Lee found that people display “’public information-seeking’ behaviors by making use of collective knowledge and/or opinions of others about music through reviews, ratings, recommendations, etc. in their music information seeking." Respondents in the study expressed needs for contextual metadata in addition to traditional bibliographic metadata.

Although the focus of these studies is not audio fingerprinting, their scope includes analysis of user behavior that adds to our understanding of how users identify and search for music. They suggest that the language used for music description and the ease and availability of resources to seek related information are key factors expressed by users, although there are a variety of ways in which people seek music and execute a search strategy. Audio fingerprinting provides a completely new way to search for music, but current studies fail to adequately investigate use cases and user behavior. Instead, they focus primarily on technical aspects of system design and implementation. In many instances these study methods will work, but they are limited by their scope and narrow study base.

Audio fingerprinting has shown to potentially address limitations of these methods, vastly improve music retrieval, and help satisfy the more desired functionalities associated with music activities. A first step in understanding the capabilities of such technology is to analyze the overall functionality of such applications and consider what early adopters are finding and communicating about audio fingerprinting. The most obvious place to find such documentation is in blogs, forums, e-mail lists, and other web-based mediums, through which users regularly communicate. The research presented in this master's paper recognizes the value of such sources and assesses them via content analysis to better understand what issues are being discussed by users of this technology and what solutions are being offered to address these issues.

RESEARCH GOALS

The predominant research question guiding this paper is: What issues are being discussed by users of audio fingerprinting technologies in non-academic web-based mediums?

Specific questions are as follows:

- What web mediums serve as grounds for discussion?
- How frequently is each audio fingerprinting issue being discussed?
- What are the authors' credentials?
- With what organizations are authors affiliated?
- What recommendations are being made for solving aforementioned issues?

METHODOLOGY

This study utilized content analysis to investigate the research questions stated above, because it allowed for an unobtrusive study of issues presented in web mediums. Published opinion and discussion could be observed without any interaction or interference from the researcher.

Convenient sampling was chosen to gather data, as this study was a preliminary research effort. This study was not meant to determine representative percentages of a larger population. Rather, this study was intended as a survey of issues present. Therefore, hard statistics were of relative unimportance.

To obtain the sample, the following process was used: Twelve searches were run through Google’s database during the week of November 22 – 29, 2008. The search terms used are presented in Table 1. These search terms were gathered from selected articles reviewed above, and were specific to the topic of audio fingerprinting. The sample keyword searches were executed with the expectation that they would yield adequate results in non-academic mediums, given the search in Google’s public directory (not Google scholar). Google was chosen as the search engine due to its flexibility of language. For example, a search for “audio fingerprint” will also include results for “audio fingerprints,” “audio fingerprinting,” as well as synonymous terms such as “acoustic fingerprint.” The hope was that the usage of flexible language would increase the return of relevant results:

Table 1: Search Terms Used to Create Data Sample for Content Analysis

	Search Terms
1	“Audio fingerprint”
2	“Audio fingerprint - blog”
3	“Audio fingerprint – Peer to Peer”
4	“Audio fingerprint - Technology”
5	“Audio fingerprint - User”
6	“Audio fingerprint - Consumer”
7	“Audio fingerprint - Issues”
8	“Audio fingerprint - Improvements”
9	“Audio fingerprint - Complaints”
10	“Audio fingerprint – Likes”
11	“Audio fingerprint - Dislikes”
12	“Audio fingerprint - Opinion”

Search results were analyzed, and the first ten results from each string meeting the following three criteria were included in the sample:

1. Authorship date of October 15, 2006 or later (This helped to maintain currency in issues. Articles of several years’ age may discuss issues that have been solved or become obsolete.)

2. Not published in an academic or trade journal

3. Not included in a previous search string (Subsequent searches often yielded some of the same results. Articles obtained in previous searches were thrown out and selection moved onward to the next eligible result.)

When one hundred and twenty search results had been obtained, numbers were assigned to each using Microsoft Excel’s random number generator. From there, the first thirty search results were chosen as the sample group and coded for analysis. Latent coding, which looks at the implicit meaning of the text, was chosen as the best tool for identifying the meaning of the text. For the purposes of this study, latent coding was determined to be more useful than manifest coding (which tallies the frequency in which words appear in the text), due to its traditionally greater validity. (Neuendorf, 2002)

An attempt was made to choose coding categories that were both exhaustive and mutually exclusive. Messages have been coded for only one level per variable. The definitions are presented below in Table 2:

Table 2: Definition of Variables

Date	Date of the article's publication
Web medium	Medium through which the article was published (Blog, Newspaper, Forum, or Listserv)
Document topic	Subject of message as intended by the author
Message function	Purpose of the message as intended by the author
Author's credentials	The authors' profession or other credentials
Author's affiliation	The type of organization as indicated by the words used to describe affiliation

The "Document topic" has been coded only from the body text of the posted message when feasible. A document can also have multiple topics if multiple topics are discussed. Many documents, being web-based, also include responses. Since authors posting to mediums that include comment or response features have a reasonable expectation that their text will create a discussion, the original message with compiled responses has been treated as one document created by the original author. Authors' credentials and affiliations have been noted. When either the authors' credentials or affiliation cannot be determined, these fields have been coded as "unknown."

After the codebook was established, two evaluators coded 20% of the messages as a representative sample with which to verify intercoder reliability. Cohen's kappa was used for all variables except "Document topic," which was calculated using a percentage agreement (see Table 3). Cohen's kappa is generally a stronger measure of intercoder reliability than simply measuring percent agreement between two coders, because it is calculated in a way that accounts for the likelihood that agreement between coders was due to chance. Since “Document topic” could be coded for more than one topic, and variable choices using Cohen’s kappa must be mutually exclusive, percentage agreement was calculated to establish intercoder reliability for this particular variable.

Table 3: Cohen’s Kappa/Percentage Agreement Calculated for Each Variable

Variable	**Intercoder reliability**
Date	100%
Web Medium	100%
Document Topic (Percent Agreement)	87%
Message Function	100%
Author’s Credentials	100%
Author’s Affiliation	100%

As the Cohen’s kappa and percentage agreement is equal or close to 100% for all six of the variables in the codebook, it shows that the codebook is a reliable analytical tool, reflecting consistent agreement between independent coders that is not simply due to chance.

RESULTS

Descriptive analysis has been carried out for each variable. Specifically, the number of messages for each category in each variable have been summed and calculated as a percentage of the whole. Further elaboration has been included to outline the significance of these results.

Table 4 shows the breakdown of documents analyzed by year of authorship. Given that only partial calendar years are represented for 2006 and 2008, the breakdown remains fairly proportionate across years.

Table 4: Total Number and Percent of Documents for Each Year

Year	**Count**	**Percentage**
2006	7	23.3%
2007	13	43.3%
2008	10	33.3%

The proportion of web mediums in which these documents were found is shown in Table 5. The largest group was “Blog” entries, accounting for over half of the total survey at 53.3%. “Newspapers” were the next largest group, accounting for 26.7% of the total. “Forums” represented the third largest group, at 16.7%. Finally, “Electronic Mailing Lists” were the least represented group, accounting for 3.3% of the total number of documents.

Table 5: Total Number and Percent of Documents in Each Web Medium

Web Medium	Count	Percentage
Blog	16	53.3%
Forum	5	16.7%
Electronic Mailing List	1	3.3%
Newspaper	8	26.7%

Blogs are a very accessible and cost-effective means of electronically documenting thoughts and opinions. Therefore it is not surprising that they account for the majority of web mediums examined by this study. Surveyed online newspapers included the New York Times and the Washington Post, as well as CNET’s newsletter. All of these publications enjoy wide readerships and are useful mediums with which to publish articles intended for a high volume of readers. Finally, electronic mailing lists and forums are useful means for soliciting help or opinions. As illustrated below, study results indicated they were generally utilized by users asking questions or initiating discussion.

Table 6 shows the breakdown of documents' message functions. Documents that posed “Questions or initiation of discussion” made up the majority at 60.0%. These documents were generally constructed in web mediums that allowed for comments and continued conversation. “Announcements” and “Complaints” came in second and third, at 20.0% and 16.7% respectively. Various announcements described forthcoming software and copyright/royalty agreements. “Complaints” differed from “Questions or initiation of discussion” in that they posed statements or rhetorical questions that were generally not intended for further conversation. Finally, only one document qualified as an advertisement, accounting for 3.3% of the total number of documents surveyed. In this particular scenario, the author advertised a fingerprinting algorithm that he had created.

Table 6: Total Number and Percent of Message Functions

Message Function	Count	Percentage
Question or Initiation of Discussion	18	60.0%
Announcement	6	20.0%
Advertisement	1	3.3%
Complaint	5	16.7%

Table 7 shows the breakdown of documents by topic. As previously mentioned, a document could discuss multiple topics. As a result, the summation of columns in Table 4 does not equal 100%. Of this sample, "Technical" topics were most widely discussed, at 43.3%. "Metadata," "User Interface," "Peer to Peer," "Copyright," "Royalties," and "Privacy" were covered fairly equally, between 23.3% and 30.0% of the time.

Table 7: Total Number and Percent of Topics for Each Document

Topic	Count	Percentage
Metadata	7	23.3%
Technical	13	43.3%
User Interface	8	26.7%
Peer to Peer	9	30.0%
Copyright	9	30.0%
Royalties	7	23.3%
Privacy	7	23.3%

"Technical" topics generally included discussion of algorithms, as well as "bugs" and other performance glitches found in current programs. These "bug discussions" were not limited to one particular program, but rather encompassed issues with several different programs. Also of note, several documents discussed the algorithm used in Shazam and other mobile phone-based fingerprinting programs. These documents reflect widespread amazement by a user base which has never enjoyed such sophisticated access to information from a mobile device.

One issue expressed in several documents was that of application speed and performance. Many users expressed concerns with the slow speed at which audio fingerprints are retrieved and compiled by various programs. Several solutions were offered to fix this issue, namely “doing more research” and “constructing better algorithms.” These solutions were not elaborated upon in the documents studied, but further investigation may yield solutions being investigated or implemented.

Finally, one document cited the issue of cross-platform and cross-application compatibility. This user discussed his disgruntlement with various audio fingerprinting applications not being available for Mac and Linux operating systems, as well as with fingerprinting metadata not being recognized consistently among various applications.

“Metadata” topics discussed issues and concerns of the file tagging and labeling processes used by current audio fingerprinting programs. These issues included both issues with the process, as well as complaints with tagging errors in the central databases. The issue of consistent metadata was hotly debated, as users discussed who was ultimately responsible for establishing tagging standards—whether the audio fingerprinting companies should be responsible—or if relying on user-generated content was the best measure. The latter option is similar to that of the popular website Wikipedia, where a community of subject enthusiasts generate their own metadata, and an element of collaborative revision enforces quality.

One document cited an issue where a user had downloaded metadata for a particular file and the returned results were in Chinese language, which the user could not understand. This was due to the only user-contributed metadata in the database being available in Chinese. Several solutions were offered for metadata-related issues found in various programs, including designating select individuals for content control, requiring record companies to submit their own metadata, and comparing metadata collections of various audio fingerprinting companies. However, the solutions posed in the examined documents have been discussed, but none have been implemented.

“User interface” topics discussed front-end issues mostly in the presentation of information, as well as the intuitiveness of interface design. User interface discussions were generally limited to discussions of specific application interfaces. For example, one user was having difficulty navigating to the “Help” documents in one program.

Surprisingly, several documents illustrated comments by users impressed with the album art features of various programs. In these, audio fingerprinting technology was not only used to download corresponding metadata for an audio file, but also corresponding album art. Users seem very impressed by this particular feature, and software companies will take notice.

“Peer to Peer” topics covered discussion related to audio fingerprinting implementation in current peer-to-peer networks and similar file-sharing programs. The documents studied merely talked about the processes used by audio fingerprinting companies to block copyrighted material, as well as advertisement of forthcoming programs to the consumer market. However, “Peer to Peer” documents presented significant tie-ins to the “Copyright” and “Royalty” sections discussed below.

“Copyright” topics discussed ways in which audio fingerprinting technology can be implemented to effectively maintain copyright control over audio files, as well as user opinions on these implementations. User opinions examined in this study’s documents ranged from favorable to strongly-opposed. One strongly-opposed user cited copyright control as the reason why he experienced difficulty finding songs readily available, free of charge.

“Royalties” discussed issues of artist compensation and how audio fingerprinting would aid or inhibit the process by which artists and record companies receive monetary compensation. Several documents discussed concerns with audio fingerprinting technology primarily aiding royalty payments for the Recording Industry Association of America (RIAA), while royalty payments for artists were being sorely neglected. Other “Royalties”-related documents served as announcements for specific partnerships between audio fingerprinting companies and those receiving services. (i.e. Audible Magic partnering with the Harry Fox Agency. The Harry Fox Agency handles licensing needs for over thirty-five thousand music publishers. This partnership would allow the Harry Fox Agency to collect property royalties for songs performed.)

Finally, “Privacy” topics included discussion of user concerns of privacy when using audio fingerprinting technologies. There was significant concern among the user documents surveyed that various parties (whether the RIAA or other copyright holders) would trace illegal downloads through use of audio fingerprinting programs. Users discussed possible implications of being caught and prosecuted, and cited privacy concerns as a significant reason for avoiding audio fingerprinting programs.

The breakdown for the different Authors' credentials is shown in Table 8. The plurality of document authors were anonymous or left no indication to their credentials and were therefore labeled as "Unknown." This accounted for 36.7% of all document authors. "Staff writers" are designated as regular writers for the publications in which they write, and accounted for 30.0% of documents surveyed. "Enthusiasts" were defined as avid users of audio fingerprinting software and accounted for 20.0% of the documents surveyed. Of lesser percentages were "Professionals" and "Students," who accounted for 10.0% and 3.3% respectively.

Table 8: Total Number and Percentage of Messages for Author Credentials

Author's Credentials	**Count**	**Percentage**
Enthusiast	6	20.0%
Professional	3	10.0%
Staff writer	9	30.0%
Student	1	3.3%
Unknown	11	36.7%

There was some correlation between author's credentials and topics discussed. For example, it seemed that staff writers were less focused on evaluating the user experience of audio fingerprinting programs and were more focused on events, market changes, and information pertaining to specific companies (i.e. an article on a recent business agreement with YouTube). It is reasonable to assume that this discrepancy is the result of job requirements. For example, a enthusiast blogger has more flexibility to discuss topics at will than a staff writer for a newspaper, whose agenda is driven more by major industry events that can grab attention.

There was also a marked difference between the discussion of royalties between

“Professional” and “Staff Writers,” versus “Enthusiasts,” “Students,” and “Unknown” writers. The former group primarily examined the topic as a public relations issue for software companies and their corporate users. The latter group primarily expressed concern that audio fingerprinting technologies would be used to enforce royalty collection, and that enforcement may not benefit the artists. The issue of user privacy was also widely discussed by the latter group, but not the former.

CONCLUSION

This master's paper reports on a content analysis of issues discussed by users of audio fingerprinting programs, as presented in documents in web-based mediums. While the scope of this study was limited, due to sample size and practical constraints, the results provide unobtrusive observation and insight into audio fingerprinting programs' user behavior and trends. Additionally, the study presents a research approach that can be used and modified to further study the usage patterns of audio fingerprinting while remaining grounded in data obtained directly from a comprehensive body of the most interested stakeholders.

This examination brought to light a number of concerns currently being debated by users. How can audio fingerprinting benefit individual users while protecting their privacy? How can developers and corporate users provide fast and accurate services that benefit the music industry at large without alienating individual users? The success of this emerging technology will depend on how the industry's major players navigate these critical questions.

The authors of these documents included industry professionals, newspaper staff writers, and consumers/users of audio fingerprinting software. There was a marked difference in topical conversation between these author types. Notable results include: Staff writers and professionals seemed less-focused on evaluating the "user experience" of audio fingerprinting programs, and focused more on events and market changes related to specific software companies. For example, in the topic of metadata, results showed that users are particularly concerned with tagging standards for metadata within fingerprinting programs. These issues were primarily covered by users and not really discussed by industry professionals and staff writers. Another example is the issue of artist royalty payments through audio fingerprinting software. Staff writers and professionals tended to discuss the mechanical and legal processes through which this happens, while users were primarily concerned with making sure artists received proper compensation. Other user concerns of note included privacy, system performance, and speed, particularly in favor of gaining more speed. One particularly striking result was the extent of user fascination towards cell phone audio fingerprinting systems and programs that procured not only file metadata, but album art as well.

As results indicate that audio fingerprinting is indeed a topic being extensively addressed on blogs, forums, newspapers, e-mail lists, and other web mediums, they prove that audio fingerprinting is demonstrably a topic worthy of further consideration and investigation. While this paper gives some baseline data as to general topics of discussion, more extensive studies will yield wider search results and gain more insight into user preferences. This will subsequently aid in finding solutions to the aforementioned issues and encourage future system design. Such studies can include both unobtrusive and more engaging means. For example, future studies can be modeled after this paper, but focus on specific e-mail lists, forums, or other sources, and draw larger samples. More engaging studies should include surveys, interviews, and other means of ascertaining what users find most important about their audio fingerprinting experiences.

This should include queries about specific interface features, as well as more broad social implications.

BIBLIOGRAPHY

Audible Magic Corporation. (2008). *About Audible Magic.* Retrieved November 28, 2008, from http://www.audiblemagic.com/company/about.asp

Bainbridge, D., Cunningham, S.J., and Downie, J.S. (2003) How people describe their music information needs: a grounded theory analysis of music queries, in *Proceedings of the International Symposium on Music Information Retrieval* (ISMIR 2003, October 2003, Baltimore, MD), 221-222.

Baluja, S. & Covell, M. (2008) Efficient wavelet-based audio fingerprinting. *Pattern Recognition* 41 (11), 3467-3480.

Barcelona Music and Technologies. (2008). *BMAT – Barcelona Music and Technologies.* Retrieved November 28, 2008 from http://www.bmat.com/vericast

Burges, C. J., Plastina, D., Platt, J. C., Renshaw, E. & Malvar, H. S. (2005) Using audio fingerprinting for duplicate detection and thumbnail generation, in *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing* 3 (ICASSP 2005, March 18-23, Philadelphia, PA), 9-12.

Cano, P. (2007). Content-based audio search from fingerprinting to semantic audio retrieval. Ph.D. dissertation. University Pompeu Fabra, Barcelona, Spain.

Cano, P., Batlle, E., Kalker, T., & Haitsma, J. (2002). A review of algorithms for audio fingerprinting, in *Proceedings of the 2002 IEEE Workshop on Multimedia Signal Processing* (December 9 – 11, St. Thomas, Virgin Islands, United States of America), 169-173.

Cunningham, S.J., Reeves, N., and Britland, M. (2003) An ethnographic study of music information seeking: Implications for the design of a music digital library, in *Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital Libraries* (JCDL 2003, May 27-31, Houston, TX), 5-16.

Downie, J.S., & Cunningham, S.J. (2002). Toward a theory of music information retrieval queries: System design implications, in Michael Fingerhut (Editor) *Proceedings of the Third International Conference on Music Information Retrieval* (ISMIR 2002, October 14-17, Paris, France), 299-300.

Einhorn, M., & Rosenblatt, B. (2005). Peer-to-peer networking and digital rights management: how market tools can solve copyright problems. *Policy Analysis* 534. Cato Institute. Retrieved October 29, 2008 from http://www.cato.org/pubs/pas/pa534.pdf.

Gracenote Inc. (2008) *Gracenote: About us.* Retrieved November 28, 2008, from http://www.gracenote.com/company_info/.

Haitsma, J., & Kalker, T. (2003) A highly robust audio fingerprinting system with an efficient search strategy. *Journal of New Music Research* 32 (2), 211-221.

Hefflinger, M. (2006) MySpace licenses Gracenote technology to block copyrighted Music. Retrieved November 28, 2008, from http://www.dmwmedia.com/news/2006/10/30/myspace-licenses-gracenote-technology-to-block-copyrighted-music

Jayant, N. S. & Noll, P. (1984). *Digital coding of waveforms: principles and applications in speech and video.* Prentice Hall, Englewood Cliffs, N.J.

Lee, J. H. & Downie, J. S. (2004) Survey of music information needs, uses, and seeking behaviours: Preliminary findings. Paper presented at ISMIR 2004: 5th International Conference on Music Information Retrieval, Barcelona, Spain.

McNab, R. J., Smith, L. A., Witten, I. H., Henderson, C. L., & Cunningham, S. J. (1996). Towards the digital music library: tune retrieval from acoustic input. (Working paper 95/29). Hamilton, New Zealand: University of Waikato, Department of Computer Science.

MediaGuide Inc. (2008). *About us – MediaGuide Inc.* Retrieved November 28, 2008, from http://www.mediaguide.com/company/index.php

Mufin GmBH. (2008) *Mufin – About us.* Retrieved November 28, 2008, from http://www.mufin.com/us/about.html

MusicBrainz Foundation. (2008) *About MusicBrainz – MusicBrainz.* Retrieved November 28, 2008 from http://musicbrainz.org/doc/AboutMusicBrainz.

MusicIP. (2008). *Music IP Mixer.* Retrieved November 28, 2008 from http://www.musicip.com/mixer/index.jsp.

Nielsen Broadcast Data Systems. (2008). *About Nielsen BDS.* Retrieved November 28, 2008 from http://www.bdsonline.com/about.html

Pandora Media, Inc. (2008). *Pandora Radio – Listen to Free Internet Radio, Find New Music.* Retrieved November 28, 2008 from http://www.pandora.com/corporate/.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks: Sage Publications.

Shazam Ltd. (2008) *Shazam on iPhone.* Retrieved November 28, 2008, from http://www.shazam.com/music/web/pages/iphone.html.

Shrestha P., and Kalker, T. (2004) Audio fingerprinting in peer-to-peer networks, in *Proceedings of the 5th International Conference on Music Information Retrieval* (ISMIR 2004, October 10-14, Barcelona Spain), Audiovisual Institute Pompeu Fabra University, 341-344.

APPENDIX

Codebook for Content Analysis:

Unit of Analysis: Non-academic online articles

Date: The date of article's publication

Web Medium:
1. Newspaper
2. Blog
3. Web forum
4. Listserv

Document Topic:
1. Metadata
2. Technical
3. User Interface
4. Peer to Peer
5. Copyright
6. Royalties
7. Privacy

Message Function:
1. Question or Initiation of Discussion
2. Announcement
3. Advertisement
4. Complaint

Author's Credentials:
1. Enthusiast
2. Professional
3. Staff Writer
4. Student
5. Unknown