Music Identification based on Audio-Fingerprinting

Yashwant singh Katailiha[1] ,Anurag Mishra[2], Vishal[3], Prajakta Naregalkar[4] Patil M.V.[5]

[1]Elex Dept. BharatiVidyapeeth Deemed to be Univ. College of Engineering Pune, Final Year
[2]Elex Dept. BharatiVidyapeeth Deemed to be Univ. College of Engineering Pune, Final Year
[3]Elex Dept. BharatiVidyapeeth Deemed to be Univ. College of Engineering Pune, Final Year
[4]Asst Prof., Elex Dept. BharatiVidyapeeth Deemed to be Univ. College of Engineering Pune
[5]Asst Prof., Elex Dept. BharatiVidyapeeth Deemed to be Univ. College of Engineering Pune

***CorrespondingAuthor**Mrs. M.V. Patil
Asst. Professor at BharatiVidhyapeethDeemed to be University College of Engineering
Pune, Maharashtra

singh.yashwant33@gmail.com
anurag.mishra0910@gmail.com[3]wishal054@gmail.com[4]prnaregalkar@bvucoep.edu.in[5]mvpatil@bvucoep.edu.in

ABSTRACT

Music identification based on audio finger printing is a digital summary condensed, a finger print generated deterministically from an audio signal, which can be used to identify an audio sample or to quickly locate similar items in the audio database. In this document, a set of audio fingerprint-based music recognition scheme is intended based on the audio fingerprint and the audio fingerprint database is built to understand the target music recognition function application. First, the audio fingerprint system encodes and decodes the audio file and then obtains the time spectrum. This technology defines a solid finger printing algorithm that can characterize a specific audio event. Within a media file and within a database identifying comparable occurrences. Audio fingerprinting systems recognize audio music material and search for recordings with the same music characteristics for a reference database.

These systems can find corresponding recordings even if the query is recorded in a public space with added noise. Different audio fingerprinting algorithms are better to identify various kinds of queries, such as brief queries or big amounts of noise in the signal. The literature contains few extensive comparisons of fingerprinting technologies that compare the precision of fingerprinting algorithms to a broad spectrum of queries.

1. INTRODUCTION

1.1 Audio-Fingerprinting

Audio fingerprinting is used to take a brief sample of an unidentified recording of audio and to collect recording metadata. This is done by turning the data-rich audio signal into a sequence of brief numerical values (or hashes) aimed at identifying a musical recording uniquely. For millions of known audio recordings, audio fingerprinting systems have big fingerprint databases. The query's fingerprint is produced to locate an unidentified audio recording query and compared to the reference database to discover recordings with identical or similar hashes of fingerprint.In the same manner as a individual can, an audio fingerprinting system should be able to recognize song recordings. If a individual can recognize a song from a brief audio clip, then the same should be possible for an optimal computer system. This implies that the perceptual elements of the data contained in the file should be used by a hashing algorithm and not just the manner the file itself is digitally coded.

1.2 Applications of fingerprinting

For a multitude of individuals who produce and consume music, audio fingerprinting is helpful. Fingerprinting systems can be used to locate unknown songs, recognize property proofs, track music as it is distributed to customers via radio or other techniques of production, or add importance to customers (Gomes et al. 2003). Consumer playback systems can inspect an audio signal's fingerprint to determine if that signal is allowed to be played. Music retailers can guarantee that they do not unknowingly duplicate audio for which a client does not have a permit to copy, for instance, bulk CD copying businesses. To generate an precise list of the audio broadcast by a radio station, broadcast monitoring can be conducted. This list can be used to guarantee proper payment of royalties to performers whose music is being performed. Batlle, Masip, and Guaus (2002) call this method "song spotting," a process that includes first separating songs from other non-musical material in an audio stream (e.g. separating songs from DJs speaking or advertising) and then performing audio fingerprints on music sections to recognize the artist and song name. Online services such as the YouTube video sharing site use a comparable method to song spotting.

1.3 The audio fingerprinting process

Most fingerprinting algorithms follow a prevalent series of steps when transforming an audio signal into a pre-processing, framing and overlapping fingerprint(Figure 1.1): transforming and analyzing, extracting features and generating fingerprints (Cano 2007). The first step, preprocessing, transforms all input signals for evaluation by the algorithm into a prevalent format. This phase often includes converting the input into a mono signal and reducing the sampling rate from the 44,100 Hz normal CD rate. For distinct algorithms, the precise method varies. Framing and overlapping determine the number of samplesThe audio signal to consider when converting to the frequency domain from the time domain. To assist in calculations, each frame has a window applied to it, and the frames are processed from the time series signal in overlapping chunks. The time-series video frames are then taken by an algorithm and converted into a frequency-domain signal from which more data can be obtained. The method of extraction of the function requires the signal that has been transformed into the domain of frequencies and selects salient characteristics used to characterize the audio. Finally, after selecting and extracting the features from the signal, they need to be converted into a fingerprint representation that can be used.

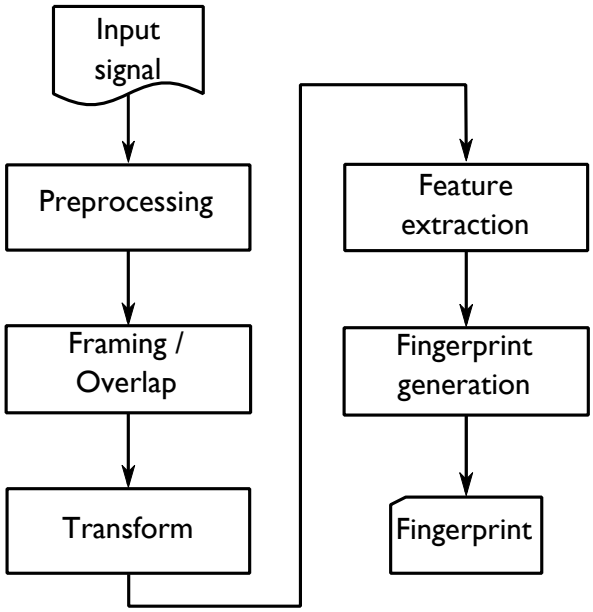


Figure 1.1 Common steps performed in audio fingerprinting algorithms to convert audio to a fingerprint.

2. LITERATURE SURVEY-

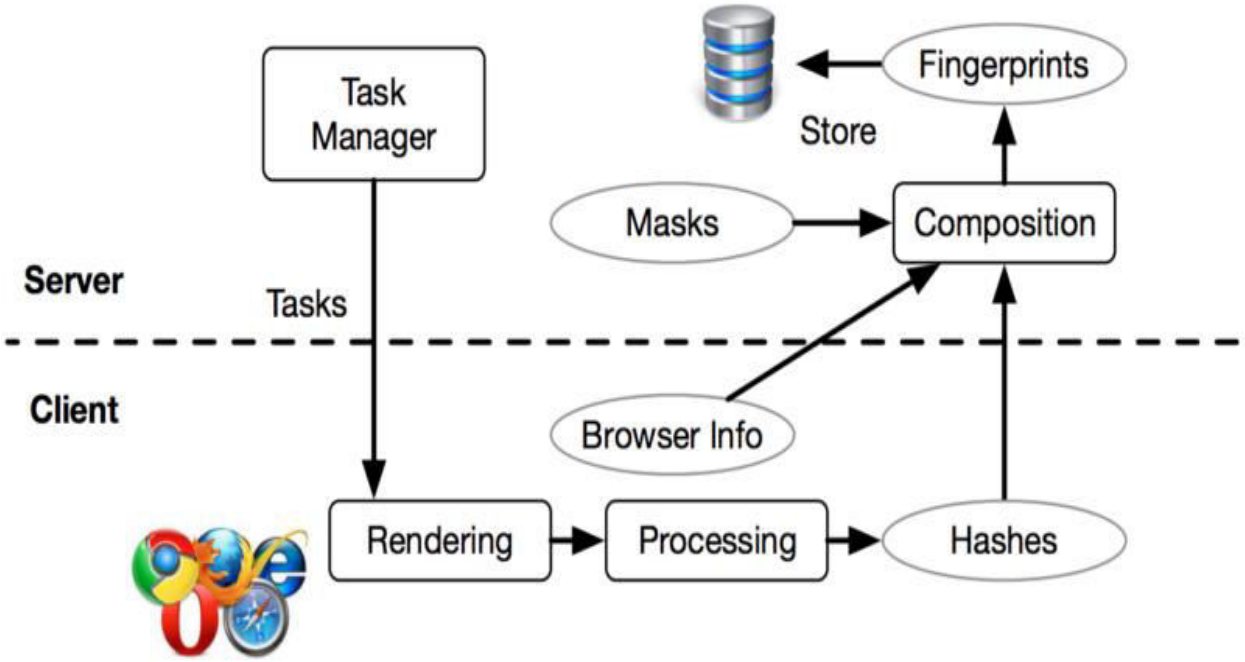
According to P.J.O Doets, M. MenorGisbert and R.L. Lagendijk[1] there are three groups into which audio fingerprinting can be categorized: Group 1: Systems that use features based on multiple subbands, namely Philips' Robust Hash algorithm, reported to be very robust against distortions. Phillips uses Haitsma&Kalker's algorithm. Group 2: Systems that use features based on a single band such as the spectral domain, namely Avery Wang's Shazam and Fraunhofer'sAudioID algorithms. Group 3: Systems using a combination of subbands or frames, which are optimized through training, namely Microsoft's Robust Audio Recognition Engine (RARE) which uses Hidden Markov Models (HMMs).

For this paper we are only interested in Group 2, as the commonly known algorithm, Avery Wang's[2]Shazam, which falls in this group, was chosen for this study. III. OPERATION All audio data used in the advertisement operation will be sampled from radio advertisements heard on your everyday radio station. Both Avery Wang's Shazam and Haitsma&Kalker's [3] algorithms were applied in the following scenario, but Avery Wang's algorithm was deemed more suitable for today's radio advertisements and was therefore studied in greater detail. Avery Wang claims that for a database of 20 000 tracks, implemented on a PC, the search time is between 5 to 500 milliseconds.
As the code is not available, adaptive code for MATLAB™ was produced by Dan Ellis [4].

Robert Macrae[5] of C4DM Queen Mary University, London, altered the code for use in the Windows environment and the authors in turn altered and implemented the code for use in advertisement identification. The code was reproduced in VB.NET. The proposed algorithm makes use of a spectrogram. The spectrogram is the squared magnitude of the STFT (Shorttime Fourier Transform). 2 (1)=ωωspectrogram t STFT t (,) (,). Byrd, D., and T. Crawford[5] has proposed the spectrogram is divided into small fragments (typically 512 points) which are called windows or frames. This is the shared basis of Group 2.

The differences between the fingerprint algorithms in the group typically involve how much the frames overlap, how the fingerprint is defined in the frame and the storing and searching of the fingerprints.
Avery Wang's and Chang, S., T. Sikora[6], Shazam algorithm uses the energy peaks in the frame and form spectral pair landmarks. They chose spectral peaks for their robustness against noise and approximate linear superposability. The local maxima within a defined section are grouped into pairs.

3. PROPOSED METHOD-In the early two thousand, the first audio fingerprinting approaches that remain relevant to this day emerged and nearly all use the same general framework. For music businesses in specific, audio fingerprinting has been and still is of great business concern to this day and there are many business apps and services (primarily mobile) that use such methods, including Gracenote2 and Shazam, among others. Usually, the



services offered allow consumers to identify unknown music tracks they hear in their daily lives and possibly boost the amount of purchases of music.

Figure 3.1 Hardware Implementation between Server and Client

With ever-increasing audio and video media databases, there is increasing demand for effective techniques of identifying, sorting, and locating comparable or identical documents. This technology defines a solid acoustic fingerprinting algorithm that can characterize a specific audio event in a media file and identify comparable occurrences in a database. In the context of rapidly identifying music and film copyright infringements, managing private or public media databases, or identifying an incident as recorded from various instruments with different background noise, the technology has excellent usefulness.

4.Working

4.1Spectogram

A spectrogram is a graphic depiction of a signal's frequency spectrum as it differs over moment. Spectrograms are sometimes referred to as sonographs, voiceprints, or voicegrams when applied to an audio signal. They may be called waterfalls when the information is depicted in a 3D matrix.Spectrograms are widely used in music, sonar, radar, voice, seismology, and other areas. Audio spectrograms can be used to phonetically distinguish spoken phrases and analyze the different animal calls.

4.1.1 Format

A popular format is a graph with two geometric sizes: one axis displays time, and the other axis reflects frequency; a third dimension is displayed by the intensity or hue of each element in the picture, showing the amplitude of a specific frequency at a specific moment.There are many format variations: the vertical and horizontal axes are sometimes switched, so time runs up and down; sometimes as a waterfall plot where the amplitude is represented by the height of a 3D surface instead of color or intensity.Depending on what the graph is used for, the frequency and amplitude components may be either linear or logarithmic. Usually, audio would be depicted with an axis of logarithmic amplitude (likely in decibels or dB), and frequency would be linear to emphasize harmonic interactions, or logarithmic to emphasize musical tonal interactions.

4.1.2 Generation

Light spectrograms can be developed over moment immediately using an optical spectrometer. Spectrograms can be produced in one of two ##s from a time-domain signal: approximated as a filterbank resulting from a sequence of band-pass filters (this was the only manner before the introduction of contemporary electronic signal processing) or calculated from the time signal using the Fourier transform. These two techniques effectively shape two distinct depictions of time-frequency, but under certain circumstances they are equal.

Usually, the technique of bandpass filters utilizes analog processing to split the input signal into frequency ranges; the magnitude of the output of each filter regulates a transducer that registers the spectrogram as a picture on paper. In the time domain, digitally sampled information is split into pieces that generally overlap, and Fourier is converted to calculate the frequency spectrum size for each piece. Then each chunk refers to a vertical line in the picture; a measurement of magnitude versus frequency for a particular moment in time (the chunk midpoint).

4.2 Shazam:-

In Shazam's situation, their algorithm selects locations where the graph has highs, marked as "greater power content." This seems to work out to approximately three points per song in reality. Focusing on audio peaks significantly decreases the effect on audio detection of background noise. Shazam creates their catalog of fingerprints as a hash table, where frequency is the core. Instead of marking a given point in the spectrogram, they label a couple of marks: the "maximum strength" plus a second "anchor point". Their database key is therefore not just a single frequency, it is a hash of both points ' values. This contributes to fewer hash collisions that in turn speed up catalog scanning by multiple magnitude orders by enabling them to take higher benefit of the steady look-up moment of the table.
This acoustic fingerprinting technique enables apps like Shazam to be able to distinguish between two strongly associated song covers.

5. PROPERTIES OF AUDIO FINGERPRINTING-

The specifications are highly dependent on the implementation but helpful for evaluating and comparing various audio fingerprinting techniques. The IFPI (International Federation of the Phonographic Industry) and the RIAA (Recording Industry Association of America) attempted to assess several recognition schemes in their Request for Information on Audio Finger Printing Technologies. Such schemes must be effective and durable in terms of computation. A more comprehensive list of demands can assist differentiate between the distinct methods:

- Accuracy:Number of correct identifications, missed identifiers and incorrect (false positives) identifications.
- Reliability: For copyright enforcement organisations, methods for evaluating whether a request is present or not in the database of products to be identified are of significant significance in the generation of play list. In such cases, even at the cost of missing actual matches, if a song has not been broadcast, it should not be identified as a match. For example, approaches to dealing with false positives were handled in. In other apps, such as instant labeling of MP3 files (see chapter on apps), avoiding false positives is not such a compulsory necessity.
- Robustness: Ability to correctly recognize an object, irrespective of the amount of transmission channel compression and distortion or interference. Ability to recognize whole films from a few seconds lengthy extracts

(known as cropping or granularity), requiring techniques to address the absence of synchronization. Pitching, equalization, background noise, D / A-A / D transformation, sound coders (like GSM and MP3), etc. are other causes of degradation.

- Security:Vulnerability of the cracking or manipulation alternative. The manipulations to cope with are intended t o confuse the fingerprint identification algorithm as opposed to the robustness necessity.Versatility: Ability to identify audio regardless of the audio format. Ability to use the same database for different applications.
- Scalability: Performance with very big title databases or a big amount of identifications at the same time. This impacts the system's precision and complexity.
- Complexity: It refers to the computational cost of extracting fingerprints, the size of the fingerprint, the search complexity, the complexity of comparing fingerprints, the cost of adding new items to the database, etc.
- Fragility: Some apps may need to detect modifications in content, such as content-integrity checking schemes. This is opposite to the robustness necessity, since the fingerprint should be solid in order to preserve content, but not other distortions.

6. Experimental Results

The assessment system's primary database and queue software is mounted on a single desktop. Each worker machine can link to this server so that outcomes can be written and job can be found. All workers have access to the same test files that can be accessed at the same time

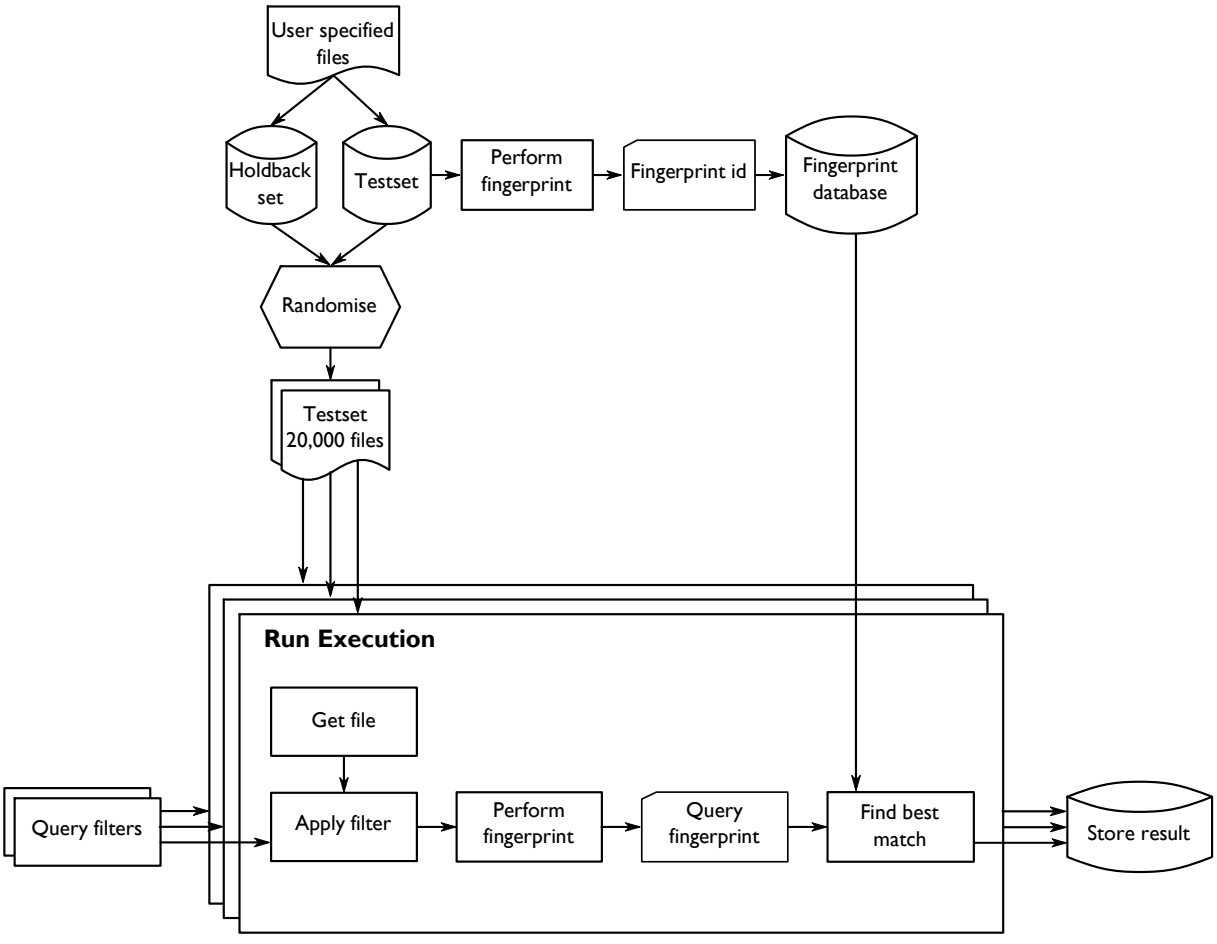


Figure 6.1 Evaluation of framework imports audio to fingerprinting algorithms

Research is a continuous process. An end of a research project is a beginning to a lot of other avenues for future work. Following aspects are identified for future research work in this area.

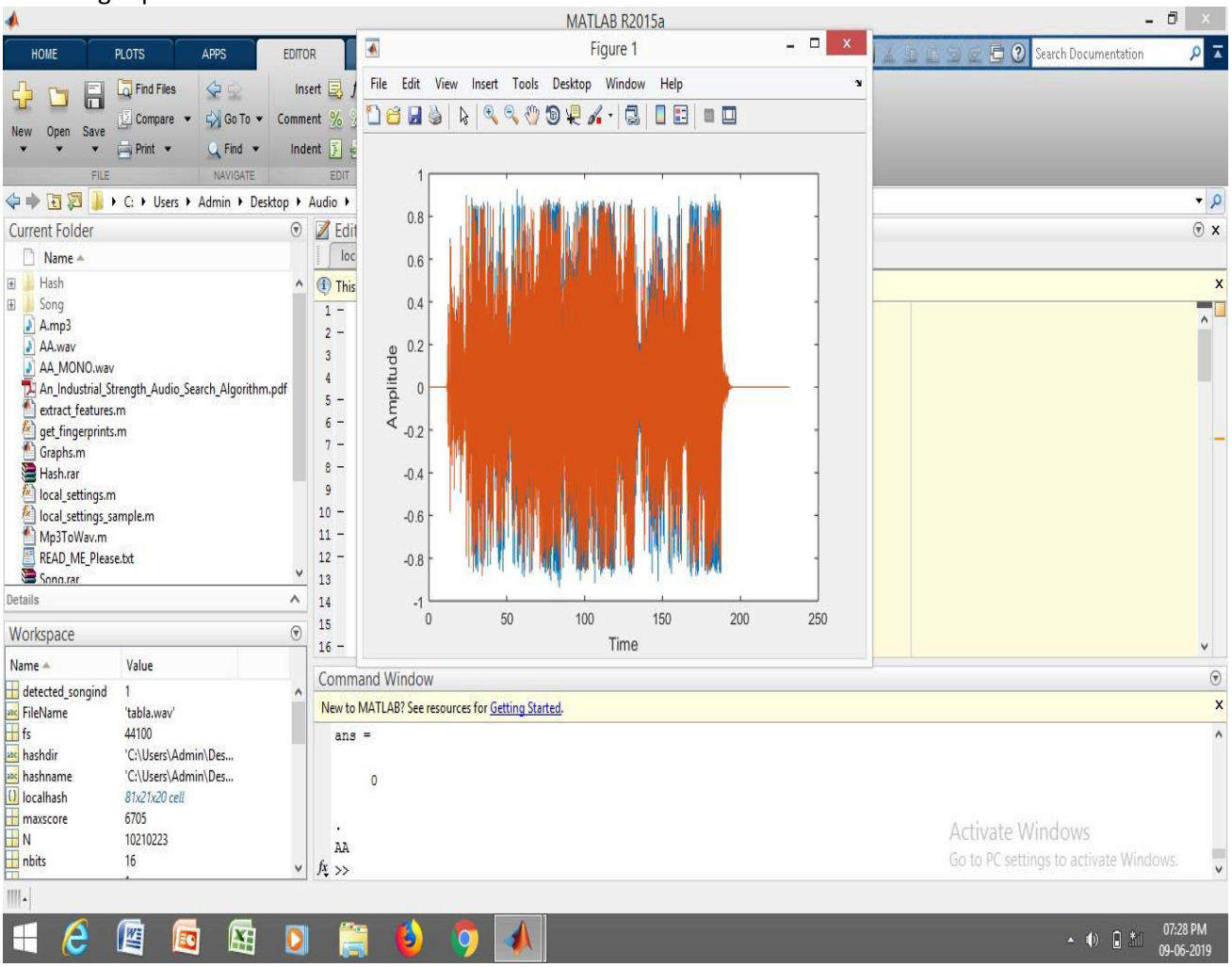


Figure 6.2 Amplitude vs Time representation of a sample of Audio signal

Research is endless. The work done in this thesis could be further researched upon and extended by considering various other sophisticated advanced simulation tools, both in the hardware and in the software levels. The hackers of the world need to come up withrobust anti acoustic fingerprinting technology. This should basically involve adding random noise without reducing listening quality. Theoretically, it should be easy to come up with.

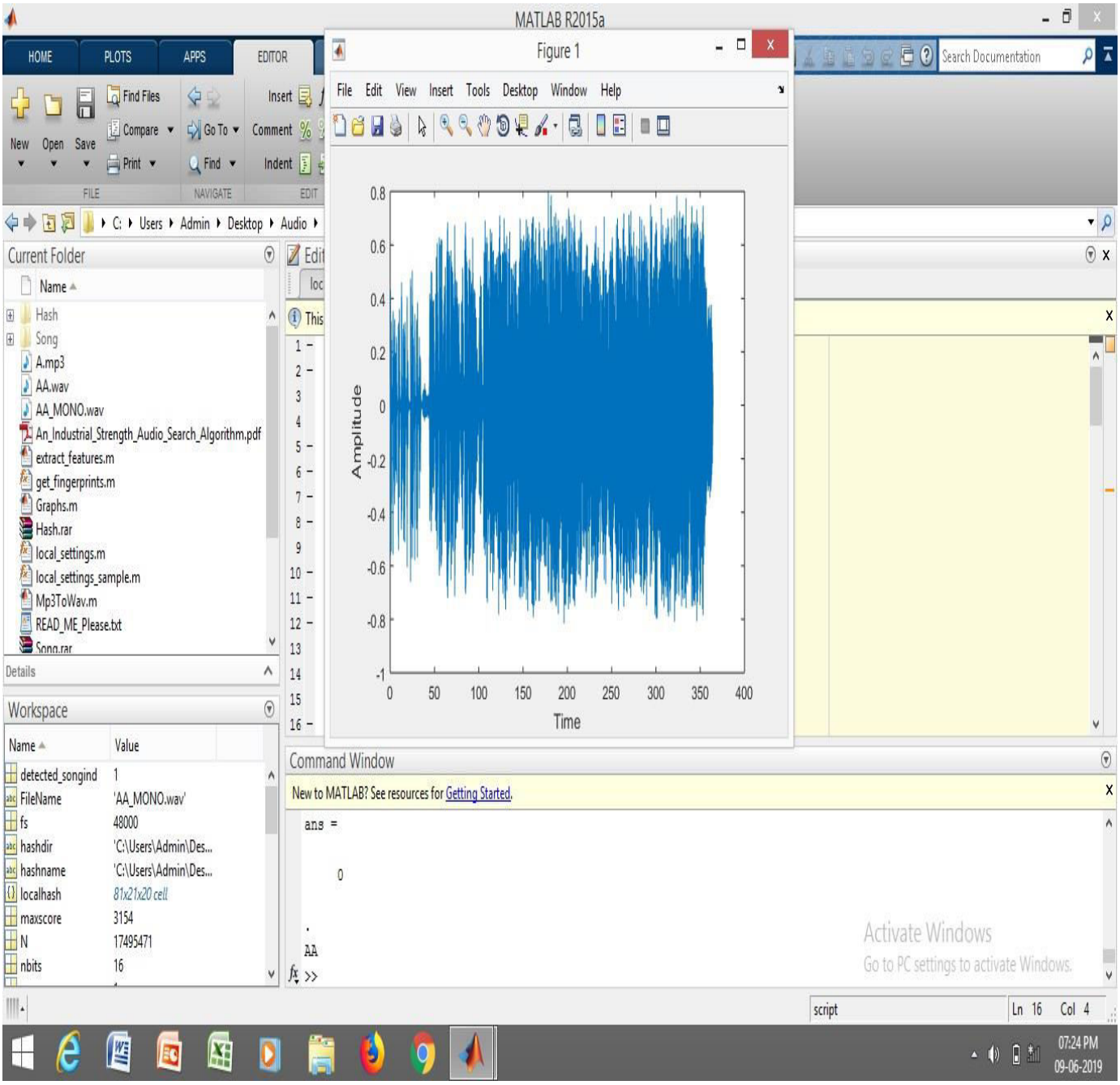


Figure 6.3 Amplitude vs time representation of another audio signal

Any info on this topic will be appreciated as it becomes available in the future. Acoustic fingerprints are generally designed to analyse audio in the same general way that humans perceive it; The worst anyone could probably pull off is to make fingerprints not match with the already existing and & quote clean & quote recordings - but the tampered copies would still be identified as one. I would say it's also economically infeasible to separately tamper each copy of a recording, with different seeds – unless produced in very low quantities. In short, sceptical about the effectiveness of this, not to mention that I fail to see any practical economical use for such a technology. Research is a continuous process. An end of a research project is a beginning to a lot of other avenues for future work.

REFERENCES:

1. M.MenorGisbert, **Reginald L. Lagendijk**: "*Security, Steganography, and Watermarking of Multimedia Contents" 2006*: 60720L

2. **Avery Wang** *is the Founder & Chief Scientist at Shazam Entertainment"*. 4, 5616 BA

3**. J., J. Haitsma, and T. Kalker**.."*Linear speed-change resilient audio fingerprinting. In Proceedings of the IEEE Workshop on Model based Processing and Coding of Audio"*2002.

4. **Dan Ellis**, "*Robust landmark-based audio fingerprinting"*2009 Last accessed 15 December 2012, http://labrosa.ee.columbia.edu/matlab/fingerprint.

5. **Byrd, D., and T. Crawford**. 2002. "*Problems of music information retrieval in the real world. Information Processing & Management"*38 (2): 249-72

6.**Chang, S., T. Sikora, and A. Purl**. 2001. "*Overview of the MPEG-7 standard. IEEE Transactions on Circuits and Systems for Video Technology"*11 (6): 688-95.