Landmark based Audio Fingerprinting for Naval Vessels

Muhammad Abdur Rehman Hashmi
Department of Electronics and Power Engineering
Pakistan Navy Engineering College, NUST-PNEC
Karachi, Pakistan
abdul.rehman2015@pnec.nust.edu.pk

Rana Hammad Raza
Department of Electronics and Power Engineering
Pakistan Navy Engineering College, NUST-PNEC
Karachi, Pakistan
hammad@pnec.nust.edu.pk

Abstract— **This paper presents a novel landmark based audio fingerprinting algorithm for matching naval vessels' acoustic signatures. The algorithm incorporates joint time - frequency based approach with parameters optimized for application to acoustic signatures of naval vessels. The technique exploits the relative time difference between neighboring frequency onsets, which is found to remain consistent in different samples originating over time from the same vessel. The algorithm has been implemented in MATLAB and trialed with real acoustic signatures of submarines. The training and test samples of submarines have been acquired from resources provided by San Francisco National Park Association [14]. Storage requirements to populate the database with 500 tracks allowing a maximum of 0.5 Million feature hashes per track remained below 1GB. On an average PC, the database hash table can be populated with feature hashes of database tracks @ 1250 hashes/second achieving conversion of 120 seconds of audio data into hashes in less than a second. Under varying attributes such as time skew, noise and sample length, the results prove algorithm robustness in identifying a correct match. Experimental results show classification rate of 94% using proposed approach which is a considerable improvement as compared to 88% achieved by [17] employing existing state of the art techniques such as Detection Envelope Modulation On Noise (DEMON) [15] and Low Frequency Analysis and Recording (LOFAR) [16].**

Keywords— Under water warfare; acoustic signatures, naval vessels; audio fingerprinting; pattern recognition

I. INTRODUCTION

The importance of potent Underwater Warfare (UWW) capability for a navy cannot be overemphasized. An Electronic Support Measure (ESM) system onboard a naval vessel monitors the electro-magnetic signatures. The ESM system detects, identifies and classifies the raw contacts into friend and foe, which thereafter are processed as per the prevalent rules of engagement. For a platform to be potent, UWW system has a similar overwhelming requirement for underwater environment. There are two types of SONARs (Sound Navigation and Ranging) active and passive. An active SONAR transmits sound energy in a known direction and calculates the time taken for the echo to arrive back. This information provides range and bearing of targets to the SONAR operator. Whereas, a passive SONAR only acquires acoustic signatures of other vessels. These raw signatures are then manually identified as valid acoustic signal (while omitting sea noise) by a human operator having adequate training on hydrophone listening. The complex part then includes categorization of this acoustic signature generating platform into submarine or a ship followed by identification of it as a friend or foe. e.g. Is it USS Shark or USS Zumwalt? Manual assessment of all this process is very complex and comes with a huge risk. Systems such as Landmark based Acoustic Target Identification System (LATIS) aid and automate this manual and cumbersome process with little to no intervention by the human operator. Using a database of saved acoustic signatures, the system autonomously queries the raw intercepted data with indexed objects in the database for a possible match thereby increasing the UWW capability onboard.

Analyzing an audio signature individually using time domain will provide signal amplitude stamps that comes with lower insight compared to frequency domain analysis. Conversely, simple frequency domain analysis becomes ineffective in the presence of factors such as noise and varying audio length etc. Some of the established algorithms for audio fingerprinting have been reported in [1], [2], [3], [4] and [5]. These techniques determine the frequency points at which maximum energy is concentrated. Different algorithms are then applied to these frequency onsets for matching the query sample with the database samples. The authors at [6] developed frequency histograms of the query and database samples and compared them, whereas [7] and [8] extracted chroma-based audio statistical features to achieve robust matching. A more recent work [9] derived binary images from samples' spectrograms. Upon finding a matching portion in binary images of query and database samples, the adjacent portions are then compared. The process is accelerated by using graphics processing unit. These algorithms are being utilized effectively in the music industry for identifying a query music clip from a large database of popular songs in few seconds. To the best of our knowledge, none of the algorithms reported and assessed have been applied towards Acoustic Signatures of Naval Vessels (ASNVs). Further, the performance evaluation of existing algorithms with ASNVs has not been reported so far.

In this paper, a novel landmark based audio fingerprinting algorithm has been designed for naval vessels. The algorithm developed in this paper, takes inspiration and guidance from the work reported by Professor Dr Dan Ellis [10]. The authors have also utilized the excerpts of MATLAB® routines provided at [10] which are originally designed for music application.

DOI 10.1109/FIT.2016.58

Rest of the paper is organized as follows. In Section II, LATIS novel algorithm is discussed followed by LATIS implementation details, comparison with state of the art existing systems and test results in Section III. Finally, Section IV contains the conclusion with possible future directions.

II. LATIS ALGORITHM

Onboard a naval vessel, the SONAR operator continuously monitors the underwater acoustic environment manually using headphones. When the operator identifies a valid acoustic signal (which may be that of another ship or a submarine etc.), he/she directs the SONAR beam (hydrophones) towards the possible direction of the target to acquire strengthened sample. This sample serves as input to LATIS (i.e. computer based system) for automated analysis. The acquired sample is compared with the held database to find any match. The match is indicated in terms of percentage. Details of the LATIS algorithm are explained below.

A spectrogram of audio signal consists of a constellation of frequency onsets. Hashes or lines are drawn between neighboring onsets and ensembles containing information of the first frequency onset f_1, the second frequency onset f_2 and the respective occurrence times t_1 and t_2 are generated. The composition of these ensembles may be altered e.g. keeping f_1 and $\Delta f=f_2-f_1$ instead of f_1 and f_2. The feature (hashes) extraction is done by applying this approach to both query and database samples. The query sample hashes are matched with hashes of each of the database samples. Instead of single domain dependency, the approach utilizes time-frequency analysis that entails robustness. The technique exploits the relative time difference between frequency onsets. The onset will remain nearly consistent for samples originating from the same naval vessel regardless of different time frames. It is due to this inherit property of the technique that it performs well even if the signal is masked with noise. The audio fingerprinting method effectively reduces the processing time by utilizing hashes rather than comparing each frequency onset. Further, depending upon signal strength, a match can be established even if a small number of hashes match since it is very rare to have same frequency onsets with same time difference co-incidently. Intuitively, comparing only frequency onsets without linked time details will lead to a large number of false positives. Flow diagram of a deterministic acoustic fingerprinting method is shown in Fig. 1. Below are some attributes that needs to be considered during the proposed approach.

A. Sample Filteration

The audio recording devices usually sample the SONAR hydrophone output at 44.1 kHz or similar. The recorded signal is passed through a digital low pass filter to remove

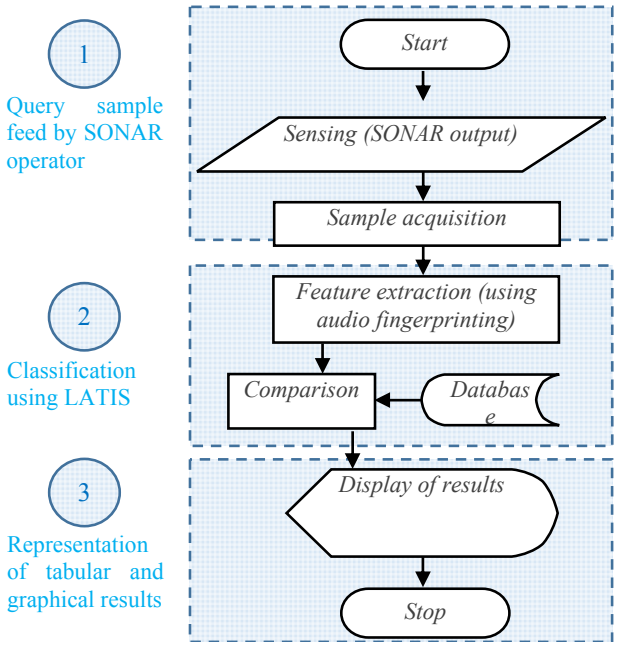


Fig. 1. Flow diagram – implementation of LATIS algorithm.

all frequency components above 5 kHz. We have used the low frequency components till 5 kHz as characteristic information of an ASNV. The database and query tracks are passed through this filter. The 5 kHz cut-off frequency has been selected since characteristic sources of an ASNV usually contain frequency components not exceeding 5 kHz. The characteristic sources are further explained in the next sub-section.

B. Selection of Sampling Frequency

During signal transformation (from time to frequency domain), the characteristic frequency components constituting an ASNV provide basis for selecting the sampling frequency. The sampling frequency is directly proportional to computational complexity and memory requirements. A sampling freqency that just retains the required frequency components is a decent option. The components of a typical ASNV [11] are listed below for readability flow:

1) Propeller and propeller cavitation noise: The formation of air bubbles near the propeller give rise to Propeller Cavitation Noise (PCN). PCN characteristics depend upon propeller speed, type of propeller and depth of propeller. The propeller noise contributes to ASNV in terms of 0.1 to 1 Hz component.

2) Machinery noise: The machinery noise is generated by vessel's engines, diesel generators and hydraulic machinery etc. These frequency components usually range upto 4 to 5 kHz.

3) Flow noise and activity noise: Flow noise is the noise caused by contact between vessel's hull and the sea.

Activity noise refers to specific activities onboard a vessel e.g. drilling and lifting etc. These noises vary from one vessel to another and also effected by the type of platform under query. For eg. some characteristic transient noises generated by a submarine includes torpedo tube opening, depth control operations and start/stop of machinery etc.

C. Generation of Hashes – Database Tracks

The spectrogram of each database track is generated. The spectrogram is a frequency domain representation using Short Time Fourier Transform (STFT) and displays the frequency onsets against time. A sampling frequency of 11 kHz is used in STFT calculations to capture all the frequency components of ASNV ranging upto 5 kHz i.a.w. Nyquist Shannon sampling theorem [12], [13]. Two frequency onsets are joined together to form a hash. The hash configuration may be selected as one of the following:

1) Hash configuration I: A hash may contain information of the first frequency component f_1 and the time at which it occurs t_1, the second frequency component f_2 and the time difference $\Delta t = t_2 - t_1$ where t_2 is the time at which the frequency component f_2 occurs. The total memory requirement depends upon the required sampling resolution. For eg. if 8 bits are used for each of f_1, f_2, t_1 and Δt, then total of 32 bits space is required for a single hash. The details of t_1 need to be ignored during comparison of hashes to make it time invariant. Later, from the hash comparison details, t_1 indicates at what instants of time the matching hashes exist in query and database samples. Hash configuration I is used in this paper.

2) Hash configuration II: A hash may contain information of the first frequency component f_1, the frequency difference between frequency f_1 and f_2 i.e. $\Delta f = f_2 - f_1$ and the time difference $\Delta t = t_2 - t_1$. Again, the total memory requirement depends upon the sampling resolution. For eg. if 8 bits are used for f_1 and 6 bits for each of Δf and Δt, then 20 bits space is required for a single hash.

Hash density is also defined during hash generation. Hash density is the number of hashes generated per unit time. Hashes are generated until the defined hash density is achieved. In order to realize desired hash density, a threshold criterion is required which defines which frequency onsets are selected first for hash generation. The threshold criterion selects the frequency onsets with largest amplitude. Once the desired hash density (e.g. 10 hashes/second) is acquired then the remaining frequency onsets are ignored.

Matching performance is directly proportional to hash density. Increasing the hash density increases the accuracy of matching but at the same time, it increases the computational complexity and memory requirement. By analyzing the algorithm performance with the same dataset and different hash densities, an optimal value can be identified.

The training and test samples of submarines have been acquired from resources provided by San Francisco National Park Association [14]. It is important to understand the nature of dataset used. ASNVs do not require to be lengthy samples. A 5 to 20 seconds long sample usually contains all the characteristic information of the platform. The other useful frequencies are those generated by the ship's propellers and machinery.

The above mentioned technique provides robustness against noise. However, the query sample may suffer distortion in terms of time skew or time scaling. Such distortion is caused by analog to digital conversion. This problem needs to be addressed by analyzing the training samples and figuring out percentage of tolerance allowed for the hash parameters. A tolerance of 1% change in the hash parameters is incorporated in this paper.

The algorithm implementation starts with initializing a Database Hash Table (DHT). Using *Hash configuration I* (32 bits per hash), allowing a space for 0.5 Million hashes per track, 500 tracks require a storage space of 954MB. The system capacity can be increased simply by creating a larger DHT requiring more space. The sampled data alongwith sampling rate of each track is acquired using 'audioread' function of MATLAB. The function accepts audio files in all popular formats i.e. wav, flac, mp3, MPEG-4 and OGG. However, MATLAB version 2005a and older versions do not have 'audioread' function. The alternate method is to use 'waveread' function and convert the track into PCM .wav format by using software like EZ CD audio converter etc. The tracks are filtered to retain components of 5 kHz and below by using a Hamming window based digital low pass filter employing the 'filter' function. Each track is resampled at 11 kHz using the 'resample' function. The spectrogram of each track is generated using the 'specgram' function. The log magnitude values of the peaks are extracted to acquire hash details until a hash density of 10 hashes/second i.a.w the threshold criterion. The hash details thus obtained are stored in the DHT already initialized. Track ID information is stored in an array to keep reference of the source of hashes in the DHT. During hash comparison, the column vector in DHT containing the occurrence time t_1 of the first frequency component f_1 is ignored to make the comparison time invariant and resistant to time skew. The track ID of the database track having maximum hashes in common with the query track indicates the match result.

3) Recording of a query sample: Upon acquiring a valid ASNV, a sample of it is recorded and saved.

4) Conversion of query sample into hashes: The query sample is converted into hashes in a similar manner as done for the database tracks. Resultantly, a set of hashes for the query sample is obtained.

5) Classification: The query sample hashes are compared with the DHT containing information on hashes of all the database tracks. How many hash matches are sufficient to declare a match safely? This number varies and depends upon iterative results obtained using training samples. In this paper, it has been found that 5 matches are sufficient to declare a match.

6) Display of Results: Match results are displayed in tabular as well as graphical form. The tabular results provide the percentage match. The percentage match does not provide the true match result since the algorithm declares a 100% match if more than 5 hashes match. It is due to the reason that the tabular output is intended to provide insight that facilitates the operator to make a decision. The graphical display shows the detailed picture since it plots and overlays the matched hashes of the query and database samples.

III. Performance Evaluation

A. Realization of LATIS

MATLAB® release R2016a on an average PC (6th generation Intel Core i5 processor @2.3 GHz, 8GB of DDR3 memory @1600 MHz, Windows 10 Home edition) has been used to implement the proposed LATIS approach. Screen shot of the system display after loading 11 in number database tracks is shown in Fig. 2. Screenshot of a tabular outcome is shown in Fig. 3, whereas graphical results are shown in Figs. 5 and 6. The query sample and the matching database sample are shown in each figure, highligting the common hashes.

B. Testing Environment

The ideal testing environment for the LATIS model is onboard a ship having a SONAR providing Hydrophone Effect (HE) output to LATIS with a dedicated LATIS operator working in co-ordination with the SONAR operator. In order to undertake offshore testing, another audio signal generating source (i.e. another PC) is used to simulate SONAR output for testing.

C. Training and Test Examples

To test the system, real acoustic signatures of some United States Navy submarines have been used. The acoustic signatures have been acquired from resources provided by San Francisco National Park Association [14]. The dataset provides signature of a submarine in varying conditions i.e. at different range, speed and combined with different noises.

For each of the vessels, two samples have been acquired recorded under different conditions of noise, depth and of varying sample lengths. Both samples originate from the same vessel. The first sample is added to the system database as the training sample. Conversion of the sample

```
Max entries per hash = 32
Target density = 10 hashes/sec
Adding #1 USN SM Adder.wav ...
Adding #2 USN SM Argonaut.wav ...
Adding #3 USN SM Bluegill 1.wav ...
Adding #4 USN SM Bluegill 2.wav ...
Adding #5 USN SM Bluegill 3.wav ...
Adding #6 USN SM Bonita.wav ...
Adding #7 USN SM Grampus.wav ...
Adding #8 USN SM Pintado.wav ...
Adding #9 USN SM Plunger.wav ...
Adding #10 USN SM Shark 3.wav ...
Adding #11 USN SM Shark 4.wav ...
added 11 tracks (119.8106 secs, 1269 hashes, 10.5917 hashes/sec)
```

Fig. 2. System display after loading database tracks.

```
RESULT  : QUERY SONAR CONTACT HAD A
   58.2000

PERCENT MATCH WITH DATABASE TRACK NUMBER
     2

NAMELY
USN SM Argonaut.wav
```

Fig. 3. Tabular display of classification result.

into hashes stored in the DHT provides necessary training to the system to recognize the source from another sample recorded under different conditions but originating from the same source. The second sample of the same vessel is used for testing. The test sample is used as a query sample to let the system find out the correct reference database sample that originates from the same source. Ninety-five such training and test samples of USN submarines have been acquired to verify system performance in terms of true positives (TP) and false negatives (FN). Conversely, an additional 95 in number test samples have been acquired which do not have reference database samples, to test system performance in terms of true negatives (TN) and false positives (FP). The samples are given arbitrary names as mentioned in Figs 2, 3, 5 and 6.

Before the testing, system must complete its training phase by populating DHT with hashes belonging to all the database tracks. System populates DHT @ 1250 hashes/ second. Eleven in number database samples comprising 120 seconds of audio, are converted into hashes in one second with a target hash density set to 10 hashes/ second.

D. Test Results

The LATIS system testing shows encouraging results. In 94 out of 95 tests, the system matched the noisy query sample with correct database sample even if query sample's length was one tenth of the database sample. Few matching patterns are shown in Figs. 5 and 6. The figures show that the algorithm is able to extract most of the characteristic hashes under conditions of time skew and noise, thus enabling correct detection and suppressing false rejection.

Chance matches are discouraged by the system due significantly low probability of having identical hash for two samples originating from different sources. While querying 95 samples having no reference in database, the system correctly rejected 85 out of 95 samples. The 10 in number false positives encountered have been further analyzed. It was found that 10 query samples although did not match the correct vessel, however, they matched to same class of submarines. So, even this information gained through false positives is not completely undesirable, since it provides closest match. Each comparison reveals that even with small length samples averaging about 8 seconds, atleast 10-15 hashes have been found common in the query and reference database sample. Thus, the system is able to perform correctly even with further shorter length samples since 5 hashes are sufficient to declare a match safely. The classification performance of LATIS is shown in terms of confusion matrix and Receiver Operating Characteristic (ROC) curve in Table I and Fig. 4 respectively. ROC curve revealed an Area Under the Curve (AUC) value of 0.9776 square units.

The existing state of the art technique for detection and classification of ASNVs is DEMON (Detection Envelope Modulation On Noise) [15] and LOFAR (Low Frequency Analysis and Recording) [16] respectively. DEMON involves narrow band analysis to find frequency components in the range of 0 to 50 Hz created by propeller cavitation noise. This in turn provides propeller characteristics including number of shafts, shaft rotation speed and propeller blade rate. LOFAR involves broadband analysis to extract frequency components contributed by ship's machinery noise. The authors at [17] employ the techniques after pre-processing the samples acquired through a passive SONAR installed on a Brazilian navy submarine. The pre-processing mainly involves Independent Component Analysis [ICA] [18] to remove signal interference. The reported classification performance is 88% on mixed signals (containing interference) pre-processed with ICA. Employment of LOFAR for target classification in [17] is inherently prune to errors. It is because varying noise in different environmental conditions may completely mask few of the characteristic frequency components, leading to false matching results. Whereas, LATIS by increasing the hash density is capable of extracting characteristic frequency components even mapped with noise. Moreover, LOFAR relies only on the frequency spectrum details to generate feature vectors and cannot distinguish between ASNVs containing same frequency components but occurring in different order. Whereas, LATIS employs a joint time – frequency technique which accurately distinguishes between ASNVs having same frequency components but occurring in different orders. Resultantly, LATIS offers a considerably better classification performance of 94% as compared to [17]. The dataset utilized by [17] is not available publicly for

TABLE I. CONFUSION MATRIX OF LATIS

Total population = 190	**Predicted YES**	**Predicted NO**
Actual YES=95	TP=94	FN=01
Actual NO=95	FP=10	TN=85

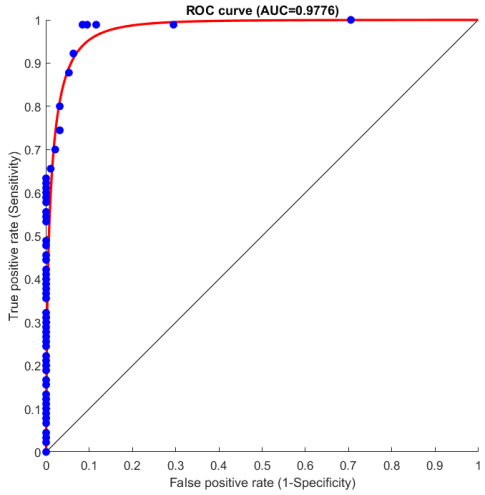
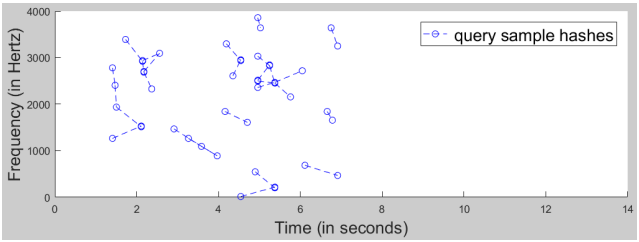


Fig. 4. Receiver operating characteristic curve of LATIS.

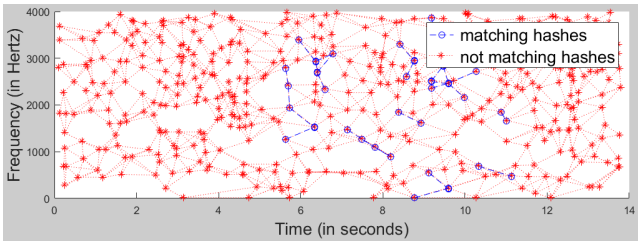
performance evaluation of LATIS. Resultantly, LATIS has been evaluated on a comparable dataset retrieved from resources provided by [14].

IV. CONCLUSION

A landmark based audio fingerprinting algorithm has been proposed to acquire promising results on acoustic signatures of naval vessels. A classification performance of 94% is achieved by employing a joint time-frequency based



(a)



(b)

Fig. 5. LATIS matching performance (a) Hashes of query audio 1 (b) Matching query hashes with SM Bluegill 1 (complete length of database sample shown).

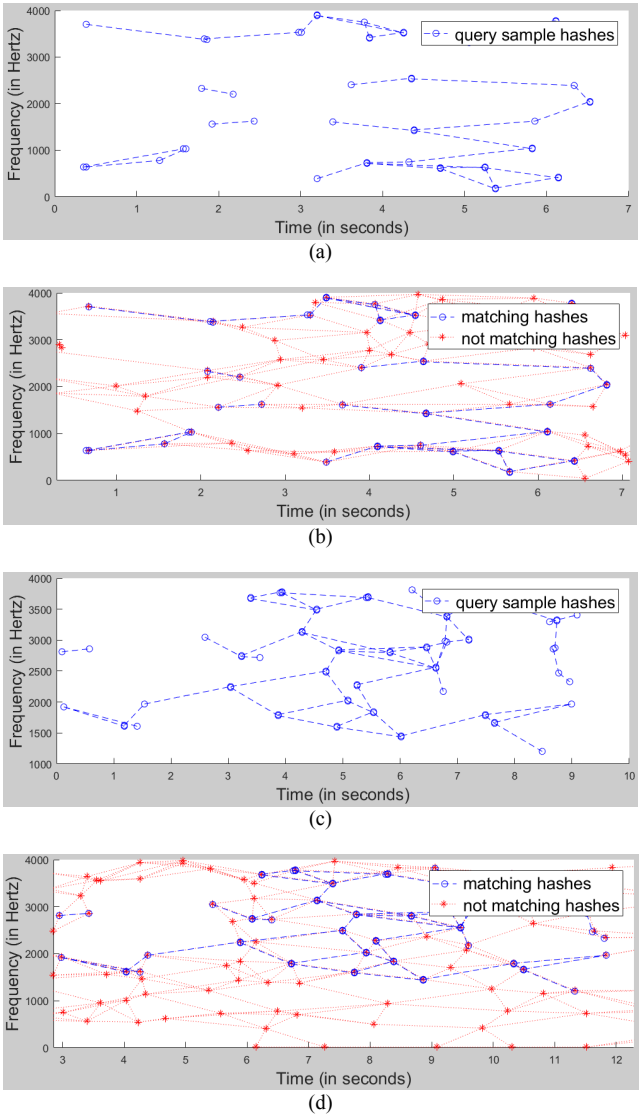


Fig. 6. LATIS matching performance (a) Hashes of query audio 2 (b) Matching query hashes with SM Bluegill 3 (c) Hashes of query audio 3 (d) Matching query hashes with SM Bonita.

technique. Same has provided a baseline for development of such models for acoustic signals. The proposed approach is considered valuable in under water detection systems. Moreover, the proposed approach may find utility in classification of other underwater acoustic sources like marine life etc.

ACKNOWLEDGEMENT

The authors would like to thank Professor Dr. Dan Ellis for sharing his work [10] and his valuable guidance during this research work.

REFERENCES

[1] P. Cano, E. Batle, T. Kalker and J. Haitsma, "A review of algorithms for audio fingerprinting," in *IEEE Workshop on Multimedia Signal Process.*, Saint Thomas, V. I., 2002, pp. 169 – 173.

[2] A. Wang and T. F. Block, "An industrial-strength audio search algorithm," in *Proc. 4th Int. Conf. on Music Inform. Retrieval*, Baltimore, Md., 2003.

[3] A. Wang, "The Shazam music recognition service", *Commun. of the ACM*, vol. 49, no. 8, pp. 44-48, Aug., 2006.

[4] H. B. Kekre, N. Bhandari, N. Nair, P. Padmanabhan and S. Bhandari, "A review of audio fingerprinting and comparison of algorithms," *Int. J. of Comput. Applicat.,* vol. 70, no. 13, pp. 24-30, May, 2013.

[5] K. Umapathy, B. Ghoraani and S. Krishnan, "Audio signal processing using time-frequency approaches: coding, classification, fingerprinting and watermarking," *EURASIP J. on Advances in Signal Process.*, vol. 2010, no.1, doi: 10.1155/2010/451695, Feb., 2010.

[6] K. Kashino, T. Kurozumi and H. Murase, "A quick search method for audio and video signals based on histogram pruning," *IEEE Trans. Multimedia*, vol. 5, no. 3, pp. 348-357, Sep., 2003.

[7] M. Muller, F. Kurth and M. Clausen, "Audio matching via chroma-based statistical features," in *IEEE Workshop on Applications of Signal Process. to Audio and Acoust.*, New Paltz, N.Y., 2005, pp. 275 – 278.

[8] F. Kurth and M. Müller, "Efficient index-based audio matching", *IEEE Trans. audio, speech, and language processing*, vol. 16, no. 2, pp. 382-395, Feb., 2008.

[9] C. Ouali and P. Dumouchel, "Fast audio fingerprinting system using GPU and a clustering-based technique," *IEEE Trans. audio, speech, and language process.*, vol. 24, no. 6, pp. 1106-1118, Jun., 2016.

[10] D. Ellis. (2009). *Robust landmark-based audio fingerprinting* [Online]. Available:http://labrosa.ee.columbia.edu/matlab/fingerprint/ (accessed 15 Feb. 2016)

[11] E. Tucholski. (2006). *United States Naval Academy SP411 Underwater Acoustic and SONAR* [Online]. Available: https://www.usna.edu/Users/physics/ejtuchol/documents/SP411/ Lesson 12.ppt (accessed 15 Feb. 2016)

[12] N. Harry, "Certain factors affecting telegraph speed," *Bell Syst. Tech. J.*, vol. 3, pp. 324–346, Apr., 1924.

[13] C. E. Shannon, "Communication in the prescence of noise," *Proc. of the IEEE*, vol. 86, no. 2, pp. 447–457, Feb., 1998.

[14] *Historic Naval Sound and Video* [Online]. Available: http://www.maritime.org/sound/ (accessed 15 Feb. 2016)

[15] R. Nielsen, *Sonar Signal Processing*, Norwood, MA, Artech House Inc., 1991.

[16] J. C. Di Martino, J. P. Haton and A. Laporte, "Lofargram line tracking by multistage decision process," in *IEEE Int. Conf. Acoust., Speech, Signal Process.*, Minneapolis, Minn., Apr., 1993.

[17] N. N. de Moura, J. M. de Seixas and R. Ramos, "Passive sonar signal detection and classification based on independent component analysis," in *Sonar Systems*, Rijeka, Croatia, InTech, 2011, ch. 5, pp. 93-104.

[18] A. Hyvarinen, J. Karhunen and E. Oja, *Independent Component Analysis*, 1st ed., New York, John Wiley & Sons Inc., 2001.