Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A NOVEL APPROACH TO GENERATE WATERMARKS USING AUDITORY FEATURES FOR AUDIO WATERMARKING

[1]TRIBHUWAN KUMAR TEWARI, [2] VIKAS SAXENA, [3] J.P GUPTA

[1]Sr. Lecturer, Department of Computer Science & Information Technology, JIIT, Noida, India-201307

[2]Asst. Prof., Department of Computer Science & Information Technology, JIIT, Noida, India-201307

[3] Ex. VC, Jaypee Institute Of Information Technology, Noida, India-201307

E-mail:tribhuwan.tewari@jiit.ac.in,vikas.saxena@jiit.ac.in , jaip.gupta@gmail.com

ABSTRACT

The increasing number of problems related to the ownership protection of the multimedia content & its authentication due to digital revolution had lead to evolution of digital watermarking techniques. Digital watermarking provides a viable solution in minimizing this problem and boosting the confidence among multimedia content i.e. text, image, video, audio providers to make available there content on a public domain. But the watermarks used in most of the watermarking methods are not personalized or unique because of which the problem of ambiguity arises if same watermark is used by two individuals. In this paper we propose a novel method to generate a unique personalized watermark which can act as authenticating information for any watermarking algorithm specially targeting an audio. The watermarks are generated through feature vectors extracted from the speech of an individual which are unique for a particular person. The uniqueness of the watermark is shown through exhaustive experiments.

Keywords: *Intellectual Property Rights (IPR), Audio Watermarking, Digital Right Management (DRM), Feature Extraction, Personalized Watermarks, Security)*

1. INTRODUCTION

The vast and frequent usage of digital multimedia data has over shadowed the use of analog multimedia data which is available to the general public. Thanks to the digital revolution. In no time a digital media can be replicated. There are a number of advantages of having a multimedia data in digital form over an analog. Some are better quality, ease in storage, ease in copying, ease of modification, ease of transformation, ease of diffusion etc. But with these increasing advantages of media availability in digital forms, the problems to counterfeit them are also increasing, leading to ownership problem like unrestricted low cost copying and unauthenticated replication and distribution etc.

In context of the low cost copying and easy replication of the digital media impacting the Indian Film Industry and Indian Economy, a study was released by U.S. India Business Council (USIBC) in March 2008. This study showed that the Indian economy is losing huge jobs and revenue as a result of piracy in India's budding entertainment industry. The data given by their study “The Effects of Counterfeiting and Piracy on India's Entertainment Industry” was alarming. The study showed as much as Rs. 16,000 crores loss each year due to piracy and 800,000 direct jobs lost as a result of theft and piracy. This is hugely afflicting India's entertainment industry **[1].** The piracy of the multimedia content has direct affect on the mindset of multimedia providers in terms of the revenue earned by them and misuse of their creations. Without any doubt we can say that more is the piracy less is the revenue. The owners and the creator put so much effort and money behind their creation but they didn’t get the desired revenues due to unlimited low cost copying & distribution. The availability of high speed communication networks operating through out the globe increases the ease of copying and their transfer from one place to the other on the internet. The digital media can be easily copied and transferred or can be exchanged with in the internet in no time. This has caused major concerns to the content provider and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

they become reluctant in providing there multimedia creation on a public domain or on the internet. Although the illegal copy transfer over the internet can be stopped by using specialized hardware which allows only the legal copy to be transferred. But the cost of implementing the hardware and also making the check through a centralized body will be too much and also this will add delay in the transfer. Rather, in order to protect the interest of the content providers, watermarking proves to be a viable solution. Based on the multimedia content whether it is image, video or audio, the watermarking technique is referred as image, video or audio watermarking. Although image & video watermarking are older techniques audio watermarking is the latest in the list of digital watermarking. One of the reason can be the complexity of the Human Auditory System (HAS) as compared to Human Visual System (HVS) and HAS not being explored fully. The watermarking algorithms discussed in the literature works for the robustness of the embedded watermark and very hardly requirement for uniqueness of watermark is given any attention. So we have come up with the idea to generate unique watermarks from the auditory features of speech which are unique for an individual.

In this paper we provide a method to generate a personalized watermark which can acts as unique authentication information for the multimedia providers. The watermark can be used in the watermarking algorithm irrespective of the media i.e. image, audio or video in which it is embedded .This is because, the personalized information is ultimately converted into a stream of bits which can be taken as a watermark for the watermarking system.

The paper is organized in the following sections. Section II provides the current state of art along with the problem of ambiguity which shows the need for a unique personalized watermark; Section III describes the auditory feature extraction process for developing the watermark, Section IV provides the method for generation of a personalized watermark Section V provides the result for uniqueness testing followed by the conclusion and future work in Section VI.

2. CURRENT STATE OF ART

The challenge in watermarking an audio media is far more than watermarking any other type of media. The reason is the wide dynamic range of audio signal as compared to the others. In addition the Human Auditory System (HAS) is far more complex & less explored than a Human Visual System (HVS). There is a universally accepted model for the HVS but this is lacking with HAS .The modeling of the HAS is dependent upon how the modeling is done for the acquisition of the sound. In case of a speech the vocal system as well as the auditory system should be appropriately modeled. The shortcoming in one can have an adverse effect on the other. The HAS perceives sounds over a wide range of frequencies from an order of Hertz to kilo Hertz. The HAS is very sensitive to additive Gaussian noise. A very small disturbance at some frequency will be audible to the ear. This implies that the HAS doesn't perceives the disturbances equally at all the frequencies. So need of a non linear system for modeling the HAS is required. Always the non linear system will be complex as compared to its linear counterpart. All the perceptual model of HAS is modeled through a bank of overlapping band pass filters with every filter working on particular frequency band. These filters are called as critical band filters. Every individual band is treated as a separate entity within the frequency spectrum. The filter bandwidth is almost constant with a value of 100 Hz up to a frequency of 500 Hz while for high frequencies it increases with the central frequency of the band [2].For the purpose of watermarking an audio, different properties of the HAS are exploited. The assessment of the sound quality after the watermarking is done is through the methods which make use of the three parameters namely loudness, critical band rate & time. Also embedding the copyright information may also involve masking the frequency of the added signal frequency. Masking of one frequency component / sound from one critical band can be done by the other frequency component with in a critical band or other critical band. The former is called as intra-band masking and the later is known as inter band masking.

The watermarking of an audio can be done in time as well as frequency domain. The different categories of watermarking techniques being discussed in the literature are:

1) Time Domain Watermarking
   - [i] Spread Spectrum based
   - [ii] Echo hiding
   - [iii] Phase coding
   - [iv] Histogram based.
2) Frequency Domain Watermarking
   - [i] Fast Fourier Transform(FFT)based

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[ii] Fractional Fourier Transform (FRFT) based
[iii] Discrete Fourier Transform (DFT) based
[iv] Discrete Cosine Transform (DCT) based
[v] Discrete Wavelet Transform (DWT) based
[vi] DCT & DWT based

3) Compressed Domain Watermarking.

The following table shows the watermarking methods along with the type of watermark used, security provided and the problem they are trying to solve.

TABLE 1 COMPARISON OF TECHNIQUES [3]

Author	Transform Used	Watermark Used	Preprocessing Of Watermark	Security	Targeted Problem
Xiangyang Wang et.al[4] 2006	DCT with DWT	Binary Image	Bit stream conversion	No	Synchronization through Barker's code
Ji Xi et.al[5] 2009	Frequency(FRFT)	Binary image	Conversion into bit Stream	No	Reduces the probability of estimating & removing watermark
Yiqingv Lin et.al[6] 2007	STFT	Coded Image watermark	Conversion into bit Stream	Yes through Multiple Scrambling	Robustness against Synchronization Attacks
Liu Xiao -li et.al[7] 2009	FFT with Psychoacoustic Model)	Pseudo number	No	No	Overcome the Characteristics of homogenization
Xiangyang Wang et.al[8] 2005	3Level DWT with DCT	Binary Image	Scrambling using Arnold Transform	No	Robustness against Synchronization Attacks
Ali Al-Haj et.al[9] 2009	2 Level DWT	Grey Scale Image	Conversion to a 2-d Matrix & then a normalized 1-D Vector	No	Appropriate Location for embedding watermark
Sartid Vongpraphip et.al[10] 2009	5 Level DWT-SVD	Binary Image	No	No	Finding the strength of the watermark through ATS
M Zamani , A B Manaf[11]	Time(LSB Modification)	Pseudo number	No	No	Low robustness of Substitution Techniques
S.T. Chen et.al[12] 2010	DWT with Entropy	Bit stream	Watermark & synchronization code forms a pseudo no.	Yes(two fold)	Entirely new prospect of watermarking is introduced

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Although a lot has been done for making the embedded watermark robust against the attacks, but a very less work has been done for generating a unique watermark. Delannay et. al. 2000 **[13]** proposed a method for secret watermark generation which are used for synchronization using 2-D cyclic patterns. They targeted image watermarking. Gang Chui et.al. 2004 **[14]** did it for a chaotic system for image watermarking. Kitanovoski et al. 2005 **[15]** proposed a method for generation of a watermark by using image content specific features. The idea was to generate the key from the robust hash which is invariant to legitimate modifications. Bibhu 2009 **[16]** used an online secret watermark generation approach. But none of the proposed methods tried to overcome the problem of uniqueness & personalization of the watermark. The uniqueness of the watermark is required so as to make watermarking the media with someone else personalized watermark difficult. If the watermarking data is not unique and personalized, there can be a situation where the watermarking process itself can be used to defame others. The simple situation is when a person X watermarks a media which belongs to person Y with the watermark of person Z. We are here assuming that the watermark of person Z can be created or generated by person X. So when illegal copy will be received it will be carrying the information of person Z although he is not the actual owner. In many of the watermarking algorithm pseudo random numbers are used as a watermark but these watermarks doesn't carry any information about the owner or the distributor itself .They are just generated through a secret key. Some of the techniques use any arbitrary image as a watermark.

The watermarking methods given in the table 1 are not using any personalized unique watermarks so they can't solve the problem described above. Also the watermark should be such that no one else can ever generate it. So we propose a frame work for the generation of the watermark. Generation of a unique watermark will act as a separate step for the embed watermark system. In the next section we will first discuss on extraction of the feature vectors from a speech or an audio signal which is used for generating unique watermark.

3. FEATURE EXTRACTION

Wavelet coefficients, Fast Fourier Transform (FFT) coefficients, Discrete Cosine Transform (DCT) coefficients, linear predictive coefficients (LPC) and Mel-frequency cepstral coefficients (MFCC) are the common features which can be used for feature vector composition of a speech signal. For the speaker as well as speech recognition systems the Mel frequency cepstral coefficients are used so often. **[17][18]** These features can be regarded as a standard in both speaker as well as speech recognition. However, from the experiments it is known that the parameterization of the MFC coefficients works best for distinguishing speakers for a speaker recognition system as compared to a speech recognition system. The error rates in detecting the correct speaker can be highly decreased if the number of MFC coefficients is increased.

In the feature extraction process the digital audio signal is basically cut into fixed size frames, i.e. the sequence of sample values, into overlapping windows of equal length. The frames are extracted out of the original signal every 10 or 20 ms. The length of a frame is longer in case it is used for speaker recognition. The length of the frame when used for this type of application is 30ms.This is done so as to increase the spectral resolution. Each frame extracted from the original speech signal in time domain is transformed to a MFCC vector. This way ultimately the original speech signal is converted into a sequence of feature vectors. Each vector represents the cepstral properties of the signal within the corresponding time frame of the window. The proposed watermark generation is described in the following section

4. UNIQUE WATERMARK GENERATION

The watermark is generated from the speech of an individual person who will be in most of the cases the multimedia owner or the distributor who has got the rights for distributing the multimedia content .The speech of the owner is recorded in a noise free environment which we are calling as the ideal environment. Prior to extraction of the auditory features from the speech a pre processing step is involved which consist of the removal of the silent period i.e. the period before the speaker starts to speak from the speech. . For boosting the signal to noise ratio (SNR) a low pass filter is used which removes some out-of-band noise. For the experiments the length of each frame is taken as 256 samples so as to make it in the powers of 2. This size helps in calculating the FFT efficiently.

The steps involved are:

Step 1: The recorded voice is taken as the input for generation of watermark.

X(n)------→ The speech signal

Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

n ------→ Total number of samples.

Step 2: Preprocessing followed by sampling done at a frequency of 22.05 kHz with frame length equal to 256 with overlapping of 50% with the adjacent frames

Step 3: Auditory Feature vectors are extracted from each individual frame.

N-------→no. of frames

size------------→the no of samples in one frame

size =256

N= length(X (n))/size*1.5

F(i) ------→ ith frame then

F(i)= X(i*1024+1 : (i+1)*1024)

i -----------→Frame_no

V1 (i) ---------→feature vector extracted from the mel frequency coefficients of the individual frame then

V1(i)=mfcc(F (i))

where i=0 to no_of _frames

V2(i)=FFT(V1(i))

The feature vector finally used for ith frame consist of the first 13 coefficients and is given by :

V (i) = V2 (i) (1:13)

Step 4: The watermark is generated by using all the feature vectors

W(x)= {(V(i)(:) } , i= 1: N

x---→ 13*N

Step 5: Reshaping the watermark into a 1-D pattern.

Step 6: Thresh holding is done to convert the generated sequence into a stream of bits.

Step 7: Step 1 to 6 are repeated for the same person and different utterances.

The thresh holding done for the generation of the bits can be used as a key to improve the security of the watermark because even if somebody uses some body else voice for watermark generation the same bit sequence can't be generated. Step 6 will yield a watermark which is not just a stream of bits but it is also carrying the information of the owner .This stream of bits is being called as a personalized watermark. The watermark now can be used as an input for the traditional watermark embedding system. When there is a dispute regarding the multimedia content the watermark will be extracted the same way it is embedded i.e. Steps 1 to 6 will be repeated. The extracted & the original watermarks will be then checked for similarity. The following section gives the result for the watermark sequence which is unique in all the cases

5. RESULT

The experiment is done by generating the watermark sequences for 10 different persons using the watermark generation algorithm. The voice is recorded for the same spoken content and the respective watermarks are generated corresponding to every person. Figure 1 shows the speech sample which shows the first 2000 samples to be of zero amplitude. These are removed by means of the low pass filter.

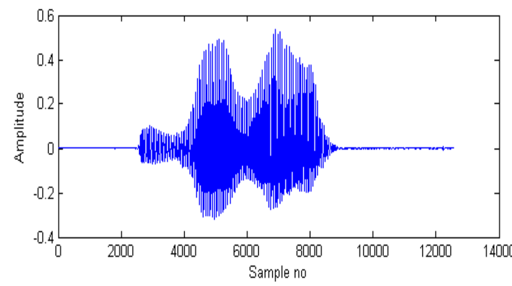

Figure 1: Speech Sample

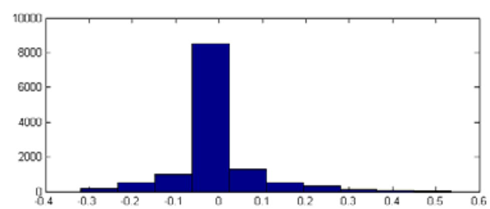

Figure 2: Histogram of the Speech Sample

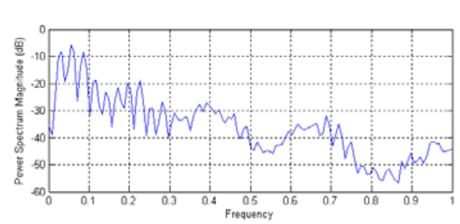

Figure 3: Power Spectral density of the Speech Sample

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

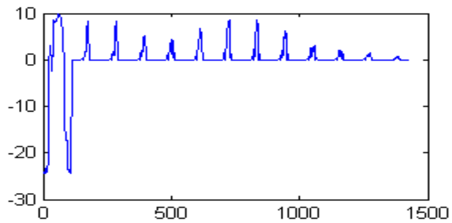

Figure 4 (a)

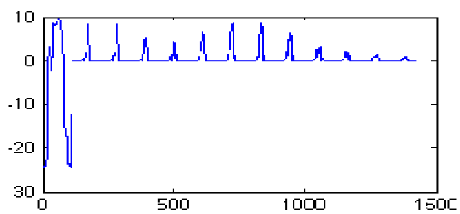

Figure 4 (b)

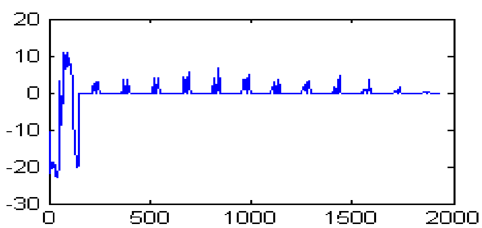

Figure4 (c)

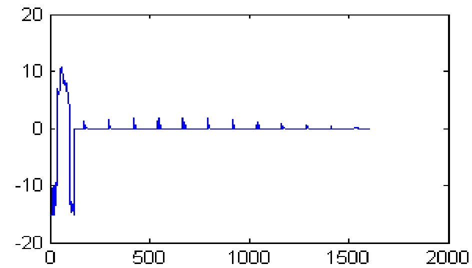

Figure. 4 (a), (b), (c), (d) Auditory Feature Vectors of 4 speech sample

Figure 2 shows the histogram of the same speech sample used for the experimentation. The peak of the histogram is at zero which shows the need of removing the silent regions of the speech which don't carry any information.

Figure 3 shows the power spectral density of a speech sample which shows that the power is distributed somewhat uniformly through out the sample. Figure 4 (a), (b), (c), (d) show the plot of feature vectors which are generated from 4 different speech samples. A matrix M is made which stores the first n elements of each feature vectors. The ith column shows the feature vector corresponding to i th sample (i.e the ith person). For showing the linear independence of two columns Gauss Jordan elimination method is used. A correlation table is made by correlating the feature vectors from each column of the matrix with itself & the other columns. For the experiment 8 samples are taken. The following table shows the result

.

Table 2: Correlation Table For 8 Feature Vectors Generated From The Speech Of Different Persons

FV-1	FV-2	FV-3	FV-4	FV-5	FV-6	FV-7	FV-8
1	0.613436	-0.10274	0.24881	0.258545	0.08903	0.475598	0.59044
0.613436	1	0.095938	0.604219	0.504819	0.319488	0.587162	0.574799
-0.10274	0.095938	1	0.270454	0.175463	0.572171	0.15933	0.008022
0.24881	0.604219	0.270454	1	0.546627	0.601178	0.506407	0.561937
0.258545	0.504819	0.175463	0.546627	1	0.540634	0.524423	0.582402
0.08903	0.319488	0.572171	0.601178	0.540634	1	0.39913	0.267451
0.475598	0.587162	0.15933	0.506407	0.524423	0.39913	1	0.596379
0.59044	0.574799	0.008022	0.561937	0.582402	0.267451	0.596379	1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In the table, entries in ith row correspond to the correlation of ith feature vector with all the other vectors which are specified by different columns of matrix M. The result shows that the highest correlation is with the same vector. This is quite obvious. The next highest correlation being 0.61 that means two feature vectors are differing by minimum of 40% which is quite acceptable. The data shows the uniqueness of the feature vector which is used for generating the watermark.

6. CONCLUSION

The paper proposes a new method to generate unique personalized watermark which can be used for any type of watermarking as the watermark is in the form of bits. It would be advantageous to embed personalized unique information which was not used by traditional watermarking methods. Work is going on for reducing the size of the auditory feature vector which can make a personalized watermark unique. This is because less is the size of the feature vector less will be the time to embed and extract the watermark. Also content based auditory features can also be used for generation of the watermark. In case of audio and video watermarking, multiple copies of the watermark generated this way can be embedded.

REFERENCES

[1] http://www.iefilmi.com/content/piracy-us-leads-1-billion-loss-indian-films.

[2] E. Ambikairajah," Auditory masking and MPEG-1 audio compression", *Electronics & communication engineering journal*, Vol. 9, No. 4, Aug 1997. pp 165-175.

[3] T.K.Tewari, V. Saxena, J.P Gupta" Audio Watermarking: Current State of Art and Future Objectives", *International Journal of Digital Content Technology and its Applications*, Vol. 5, No. 7, July 2011, pp 306- 314.

[4] Xiangyang Wang, Wei Qi, Panpan Niu, "A New Adaptive Digital Audio Watermarking Based on Support Vector Regression", *IEEE Transactions on Audio, Speech, and Language Processing*, Volume 15, Issue 8, Nov. 2007 pp 2270 – 2277.

[5] Ji Xi,"A Scheme of Robust Audio Watermarking Based on FRFT for Copyright Protection ",*In proceedings of sixth international conference on Management and Service Scienc*,(MASS) 2009, pp 1-4.

[6] Lin Yiqing , W H Abdulla, "A secure and robust audio watermarking scheme using multiple scrambling and adaptive synchronization", *In proceedings of sixth international conference on Information, Communications and Signal Processing* (ICICS),2007 pp 1 – 5.

[7] Liu Xiao-li, Sun Shui-fa, Zhu Jian-wei, Jiang Ming, "An Audio Watermarking Method of Resistance Statistics Attack Based on Psychoacoustic Model", *In proceedings of fifth international conference on Information Assurance and Security* (IAS), 2009,Vol. 1, pp. 737 – 739.

[8] Xiang Yang Wang, Hong Zhao, "A Novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT", *IEEE Transactions on Signal Processing*, Volume 54, Issue 12, Dec. 2006 pp 4835 – 4840.

[9] A Al-Haj, , L Bata, A Mohammad, "Audio Watermarking using Wavelets", *In proceedings of first international conference on Networked Digital Technologies* ,(NDT), 2009, pp. 398 – 403.

[10] S. Vongpraphip., M. Ketcham, "An Intelligence Audio Watermarking Based on DWT- SVD Using ATS", *In Proceedings of the 2009 WRI Global Congress on Intelligent Systems* (GCIS) 2009, Vol. 3, pp. 150 – 154.

[11] M Zamani, A B Manaf, "P. Magalingam., "A Novel Approach for Audio Watermarking", *Proceedings of International Conference on Information Assurance & Security* (IAS), Kualalumpur (Malaysia), 2009,Vol. 2, pp. 83 – 86

[12] S.T. Chen, "Energy-proportion based scheme for audio watermarking", *IET Journal of Signal Processing* 2010, Vol.4, No.5, pp 576–587.

[13] Delannay, D.; Macq, B.; "Generalized 2-D cyclic patterns for secret watermark generation ",*International Conference on Image Processing*, (ICIP), 2000,Vol. 2 ,pp. 77 - 79

[14] Shao Gang-chui; Zhang Min-ru;" Novel fragile authentication watermark based on chaotic system", *IEEE International Symposium on Industrial Electronics*,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2004 ,Volume: 2 , Page(s): 1491 - 1494 vol. 2 .

[15] V.Kitanovski, D. Taskovski, S. Bogdanova, " Watermark Generation using Image-dependent Key for Image Authentication", *International Conference on Computer as a Tool,(EUROCON)*, 2005. Vol. 2 pp. 947 – 950.

[16] V .Bibhu ," Online Secret Watermark Generation For Audio Files", *International Journal of Computer & Electrical Engineering, (IJCEE)*, Vol. 1, No. 5, 2009,pp. 562-565.

[17] Qi Li, Yan Huang, "Robust Speaker identification using an Auditory based Feature", *In proceedings of IEEE International conference on Acoustics, Speech & Signal Processing*, (ICASSP), 2010, pp. 4514 -4517.

[18] Hesam Tolba et.al," Auditory-based acoustic distinctive features and spectral cues for automatic speech recognition using a multi-stream paradigm". *In proceedings of IEEE International conference on Acoustics*, *Speech & Signal Processing*, (ICASSP), 2002, pp. 837-840.

AUTHOR PROFILE:

Tribhuwan Kumar Tewari received his M Tech. degree in Information & Communication Technology from JIITU,Noida ,in 2007. He is pursuing Ph.D and currently he is a Senior Lecturer at Jaypee Institute of Information Technology, Noida. His interests includes Digital Copy Right Management, Multimedia Security, Audio Processing, Object Oriented Technologies and Data Structures.