A SYSTEMATIC APPROACH TO AUTHENTICATE SONG SIGNAL WITHOUT DISTORTION OF GRANULARITY OF AUDIBLE INFORMATION (ASSDGAI)

Uttam Kr. Mondal[1] and J.K.Mandal[2]

[1]Dept. of CSE & IT, College of Engg. & Management, Kolaghat, W.B, India
`uttam_ku_82@yahoo.co.in`
[2]Dept. of CSE, University of Kalyani, Nadia (W.B.), India
`jkm.cse@gmail.com`

ABSTRACT

The major concern to authenticate an audio signal is restoring its basic audible properties and providing such a way that may not produce any unusual perception to listeners. Therefore, authenticating an audio signal needs to embed as less as minimum security code within tolerable range of perception. In this paper, perceptual encoding has been applied for structuring song signal in a specified manner with less modification of its amplitude values and followed by embedding formatted data in skewed pattern that will use for verifying any type of alternation by comparing its own structural format and systematic structure of song signal with neighbouring sampled values. A comparative study has been made with similar existing techniques and experimental results are also supported with mathematical formula based on Microsoft WAVE (".wav") stereo sound file.

KEYWORDS

Embedding Secret Message in Quality Song, Perceptual Coding, Song Authentication, Tolerance Level of Embedding Message, Bit Level Encoding.

1. INTRODUCTION

Adoption of editing and mixing technology in music – creativity is now under questions. Because, technology is grabbing the quality of voice/ music and tuning is made automatically. As a result, a new version of previously created song is generated which may be more rhythmic than original [1][2]. On the other hand, important portions of songs are being copied or added with other songs which enriched the new versions but changing the glimpse of old versions. Thus, newly created music became older before its time; industries as well as concern personals are losing revenue. Therefore, a software based sophisticated technique is required for justifying copied/ modified portions of song with some unique features and the pirated versions are to be detected as well. Another crucial thing that the adding authenticated criteria may not alter audible quality of authenticated song. Therefore, impurities/ added data along with the authenticated method need to be justified whether it is directly or indirectly affecting the audible quality of song signal [4]. Another point needs to be noted that the consistency in changing environments (in the changing values of controlling parameters like bit rate (kbps), audio sample size (bit), audio sample rate (kHz), amplitude compression, echo addition, all- pass filtering, equalization, noise addition, etc.) or format (under other compression techniques).A systematic approach with

considering all above criteria has to be incorporated to find a technique for authenticating an audio signal.

In this paper, perceptual encoding has been applied over the sampled values of song signal for representing as values in the set of discrete pre-calculated values and followed by fabricating a secret key which, is together with song signal, represents an authenticate unique signature for identifying the song from collection of similar songs. A set of experimental results has been represented to analysis its audible quality after embedding authenticating information in specified positions.

Organization of the paper is as follows. Encoding of sampled values of song signal is described in the section 2.1. Constructing secret code is presented in section 2.2. Embedding secret data is represented in section 2.3. Determine constancy in changing environment/ format is performed in section 2.4. Authentication technique is discussed in the section 2.5. The extraction is shown in section 2.6. Experimental results are given in section 3. Conclusions are drawn in section 4. References are given at end.

2. THE TECHNIQUE

The technique fabricates the authenticating signature of song signal by structuring its fraction values with the help of perceptual encoding followed by designing a secret key comparing two channel data of sampled values. The encoding technique in fraction value is limited to nearest feature line (NFL) constructed via generating a subspace with an approximate nearest value of selected fraction value [9]. Algorithms termed as ASSDGAI-ESV, ASSDGAI –FSSD and ASSDGAI –ESD are proposed as multiple levels security measure, the details of which are given in section 2.1, section 2.2 and section 2.3 respectively.

2.1. Lossy Encoding of Sampled Values (ASSDGAI – ESV)

Constructing upper boundary of sampled values of song signal with assigning near round off values up to a certain level as permitted basic hardware capability (common for available system) is used for the present lossy encoding technique [10]. The encoding technique is rounding off the value of fraction part of sampled values considering its near ceiling boundary values (determined by precision level of computing). Let, the precision level is 16. Therefore, $\frac{1}{16}$ value will represent the step value. The encoding technique is described as follows.

Step 1: As the step difference is $\frac{1}{16}$. Therefore, ceiling values can be represented by 0, $\frac{1}{16}$, $\frac{2}{16}$, $\frac{3}{16}$, …, $\frac{15}{16}$. If we allow another sublevel in a step value, then total ceiling values will be as follows 0, $\frac{0.5}{16}$, $\frac{1}{16}$, $\frac{1.5}{16}$, …, $\frac{15.5}{16}$.

Step 2: Find fraction part of sampled values of song signal for a particular position, and compare with nearest ceiling value, by determine minimum difference and substitute it by the respective ceiling value.

Step 3: Repeat step 2 for all sampled values of song signal.

For a particular application, if more precision levels are required, then find the equal difference smallest step of ceiling values, and continue the above process.

2.2. Forming Secret Systematic Data (ASSDGAI – FSSD)

As perceptual coding has been applied over the sampled values of song signal (described in section 2.1) for systematic representation of amplitude values of two channels of stereo type song, select a particular portion from these systematic amplitude values of sampled data and make the values of two channels exactly same by adjusting a particular channel's data with alternate channel for a particular region of sampled values and the extra added/ substituted values are kept in two dimensional array i.e. if the two dimensional array is added with the particular channel's data, then the data of both channels will be same for the selected portion. Hence, the array's values are substituted/ added from systematic representation of data. Therefore, this extra data also represent systematic values generated as a result of perceptual encoding which already applied over the sampled values of song signal. The array data is now used as a secret data for authenticating songs as the data is added sequentially with the particular region then, amplitude values of both channels will be same.

Therefore, embedding the extra data in specified way with avoiding any type of distortion in audible quality, secret information for authenticating song signal is to be supplied along with the song signal. Now, extracting the data and adding with the particular channel of the particular set of sampled values, a verification criteria will be generated as the sampled data of both channels will be same for this selected region.

2.3. Embedding Secret Data (ASSDGAI – ESD)

The characteristics of the sampled data of two channels have been shown in [1]-[2]. In these publications, it is proven by empirical as well as experimental result, shown that basic different of sampled values is as minimum as ignorable. Even, it may be ignore without alternation of song signal's audible quality. A set of experimental results are also shown for justifying this observation. As a conclusion, adding these deducted data in different part of sampled values will not exactly affect the audible quality of song signal as it is not alternating the audible quality when deducted from song signal.

Now, choose a number of positions from sample values, and this number will be at least 4 times of extra data array size. For example, say 100th, 200th, 300th, …, etc positions of sampled data. Embedding hidden data without losing audible quality is shown in following steps.

Step 1: Apply FFT to find frequency component of song signal.
Step 2: Make 0s in all the last consecutive 4 positions from LSB of magnitude values of song signal for the above said positions (100th, 200th, 300th, etc).
Step 3: Represent the values of hidden data in 4 parts. Say, a particular value is 0.0012; represent it by 0, 0, 1, 2 (in binary 0000, 0000, 0001, 0010 for the positions of consecutive 4 bits of LSB of the above particular positions) and store then into another array of 4 times size of extra data. Say, the new array name is S(r), whereas r is the number of divided extra data.
Step 4: Add S(r) values with selected positions of magnitude values as represented in the following expression.

SONG (100, c) = SONG (100, c) $\oplus$ S (1)
SONG (101, c) = SONG (101, c) $\oplus$ S (2)
SONG (102, c) = SONG (102, c) $\oplus$ S (3)
SONG (103, c) = SONG (103, c) $\oplus$ S (4)
…..
SONG (p $\times$ 100, c) = SONG (p $\times$ 100, c) $\oplus$ S (p)
SONG (p $\times$ 100+1, c) = SONG (p $\times$ 100+1, c) $\oplus$ S (p+1)
SONG (p $\times$ 100+2, c) = SONG (p $\times$ 100+2, c) $\oplus$ S (p+2)

SONG (p × 100+3, c) = SONG (p × 100+3, c) ⊕ S (p+3)

…..

Whereas ⊕ is the binary XOR operation between binary pattern of S(r) array's values and last 4 bits of LSB side of magnitude values of song signal SONG(n,c), c value may be 1 or 2, for representing channel number.

Step 5: Apply inverse FFT to get back the sampled values of modified song signal.

In the case of mono type song signal, two consecutive regions of sampled values may be considered for substituting the comparison of two channel sampled values.

2.4. Determine Constancy in Changing Environment/ Format

Statistical analysis has to be performed before embedding secret message, i.e. the existence of embedded code with the selected portion of song signal need to be considered after applying higher level compression methods. For example, lower levels as well as higher level frequencies are least bother for quality assurance. The audible range of the human ear is 20 Hz (0.02 kHz) to 20,000 Hz (20 kHz). The upper limit reduces with age; most adults are unable to hear above 16 kHz. The Tones between 4 and 16 Hz can be perceived via the body's sense of touch. Frequency resolution of the ear is 3.6 Hz within the octave of 1,000–2,000 Hz [8].

Therefore, if we able to send encoded message in less changed portion/region – secret message may exist in changing environment/format. The crucial portion of frequency range may be a good opt for sending the secret message as described in section 2.3. Even, trade off ratio between original song and embedded impurities (encoded message/ secret code) needs to be estimated for maintaining the tolerance level of song signal. Channel coding theorem for error correcting may be applied as this purpose [7].

2.5. Authentication

Estimated ceiling value of fraction part of sampled values and symmetric representation of secret information in bit level for a particular region of magnitude values for both channels of stereo type song creates a secure code that will use to identify the original song. Extracting embedded information and adding with selected region of particular channel a new authenticated song signal will be generated that will carry same sampled values for both channels in the particular region. The size of embedding data will be increased or decreased as required or based upon the acceptance trade-off ratio of original song and impurities (alternated bit values in frequency domain) [7].

Therefore, if any changes during processing, it will create a difference with the authenticating codes that present in the selected region of the song signal and changing a position will create difference with the hidden code in that region as well as relationship of rounding ceiling values of fraction part of sampled values also reflect the volume of changes over original song.

2.6. Extraction

The decoding is performed using similar calculations as encoding technique. The algorithm of the same is given below.

Algorithm:

Input: Modified song signal with embedded authenticating code in selected region.

Output: Original song signal.

Method: The details of extraction of original song signal are given below.

Step 1: Find bit level representation for all sampled values of selected song signal for the particular encoded region.

Step 2: Recover the alternated bit positions comparing with the information in selected bit positions of particular magnitude values.

Step 3: Apply step 2 for the all positions of the selected region of song signal.

Step 4: Deduct the extra hidden secure code from magnitude values of song signal.

3. EXPERIMENTAL RESULTS

Encoding and decoding technique have been applied over 10 seconds recorded song, the song is represented by complete procedure along with results in each intermediate step has been outlined in subsections 3.1.1 to 3.1.4. The results are discussed in two sections out of which 3.1 deals with result associated with ASSDGAI and that of 3.2 gives a comparison with existing techniques.

3.1 Results

For experimental observations, strip of 10 seconds classical song ('100 Miles From Memphis', sang by Sheryl Crow) has been taken. The sampled value of the song is given in table 1 as a two dimensional matrix. Figure 1 shows amplitude-time graph of the original signal. ASSDGAI is applied on this signal and as a first step of the procedure which is performed ASSDGAI over input song signal. The output generated in the process is shown in figure 2. Figure 3 shows the difference of frequency ratio of original and modified song after embedding secret code. From figure 3, it is seen that the deviation is very less and there will not affect the quality of the song at all.

3.1.1 Original Recorded Song Signal (10 seconds)

The values for sampled data array $x(n,2)$ from the original song is given in table 1. Whereas the graphical representation of the original song, considering all rows (441000) of $x(n,2)$ is given in the figure 1.

Table1. Sampled data array $x(n,2)$.

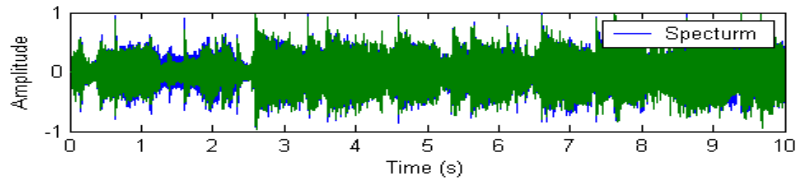| Sl no	x(k,1)	x(k,2)
…	…	…
	0	0.0001
	0.0000	0.0000
	-0.0009	-0.0009
	-0.0006	-0.0007
	-0.0012	-0.0012
	-0.0014	-0.0014
	-0.0016	-0.0017
	-0.0023	-0.0022
	-0.0027	-0.0027
	-0.0022	-0.0021
…	…	…

Figure 1. Original song ('100 Miles From Memphis', sang by Sheryl Crow)

3.1.2 Modified song after embedding secret code and rounding off the fraction part of sampled values (10 seconds)

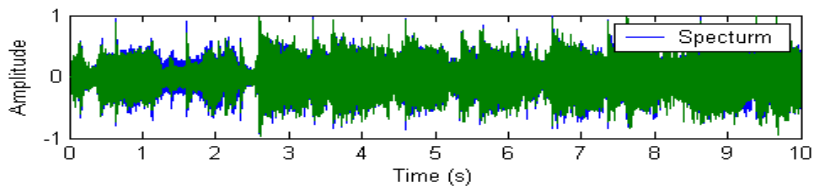The graphical representation of the modified song signal is shown in the figure 2.

Figure 2. Modified song with secure code

3.1.3 The difference of magnitude values between original and modified signals

The graphical representation of the magnitude differences of original and modified songs is shown in the figure 3.
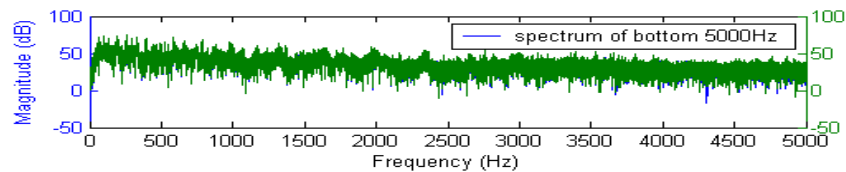

Figure 3. The difference of magnitude values between signals shown in figure 1 and 2.

3.1.4 Estimating limit of embedded code

Estimating limit of embedding data over song signal is one of the major issues when quality is a factor. For this purpose, an approach has been made for estimating the boundary of adding impurities without compromising its audible quality is done with the help of channel coding theorem in modified form as follows [7]. Channel capacity can be expressed by following formula(1)

$$C = W \log_2 (1 + \frac{P}{N_0 W})$$
(1)

bits per second (here, magnitude values per sampled set of song signal) Assigning the values of above variables with considering Shannon's limit are given as follows   W= 20,000 (approx) [considering audible range 20-20,000 Hz], C=44,100 [16-bit stereo audio signals sampled at 44.1 kHz], $P = E[X_k^2]$ whereas X_k = frequency values of original song, K=1, 2, 3, .., L

[L= length of song signal] Putting above values in the equation (1), we can easily find the noise value N_0. Spectrum density = $N_0/2$ [maximum noise] and (N-k) number of added impurities (noise) in sampled values of song signal. Where r<=C, according to channel coding theorem. Applying equation (1) to above song [16-bit stereo type sampled at 44.1 kHz] , hidden extra data (noise) that added for authenticating original song is very less than the maximum noise (0.0739), will not affect over all song audible quality. Because, only about 4000 positions have been altered from 441000 sampled values of taken song signal.

3.2 Comparison with existing systems

Various algorithms [5] are available for embedding information with audio signals. They usually do not care about the quality of audio but we are enforcing our authentication technique without changing the quality of song. A comparison study of properties of our proposed method with Data hiding via phase manipulation of audio signals(DHPMA)[3] and An Analytical Approach to Generate Unique Song Signal (AUSS) [10] before and after embedding secret message/modifying parts of signal (16-bit stereo audio signals sampled at 44.1 kHz.) is given in table 2, table3 and table4. The AD, NAD, MOS and PSNR are often used to assess the quality measurement between the original and a modified song. The higher the PSNR represents the better the quality of the modified song. Thus from our experimental results of benchmarking parameters (NAD, MSE, NMSE, SNR and PSNR) in proposed method obtain better performances without affecting the audio quality of song. Table 3 gives the experimental results in terms of SNR (Signal to Noise Ratio) and PSNR (Peak signal to Noise Ratio). Table 4 represents comparative values of Normalized Cross-Correlation (NC) and Correlation Quality (QC) of proposed algorithm with DHPMA and AUSS. The Table 5 shows PSNR, SNR, BER (Bit Error Rate) and MOS (Mean opinion score) values for the proposed algorithm. Here all the BER values are 0. The figure 4 summarizes the results of this experimental test. It shows this algorithm's performance is stable for different types of audio signals.

Table 2. Metric for different distortions

Sl No	Statistical parameters for differential distortion	Value using AUSS [10]	Value using ASSDGAI	Value using DHPMA [3]
1	MD	0.0154	0.0023	3.6621e-004
2	AD	0.0102	6.2065e-005	2.0886e-005
3	NAD	0.0101	5.3710e-004	0.0063
4	MSE	5.3738e-006	1.5073e-008	1.4671e-009
5	NMSE	2.2367e+004	1.5478e+006	8.4137e-005

Table 3. SNR and PSNR

Sl No	Statistical parameters for differential distortions	Value using AUSS [10]	Value using ASSDGAI	Value using DHPMA [3]
1	Signal to Noise Ratio (SNR)	38.0041	36.5602	40.7501
2	Peak Signal to Noise Ratio (PSNR)	54.407	58.3401	45.4226

Table 4. Representing NC and QC

Sl No	Statistical parameters for correlation distortion	Value using AUSS [10]	Value using ASSDGAI	Value using DHPMA [3]
1	Normalised Cross-Correlation(NC)	1	1	1
2	Correlation Quality (QC)	-0.0165	-0.2238	-0.5038

Table 5. Showing SNR, PSNR BER, MOS

Audio (10s)	SNR	PSNR	BER	MOS
Song1	36.5602	58.3401	0	5
Song2	35.5061	57.459	0	5
Song3	36.0677	55.8802	0	5
Song4	44.334	66.8703	0	5

The ITU-R Rec. 500 quality rating is perfectly suited for this task, as it gives a quality rating on a scale of 1 to 5 [6]. Table 6 shows the rating scale, along with the quality level being represented.

Table 6. Quality rating scale

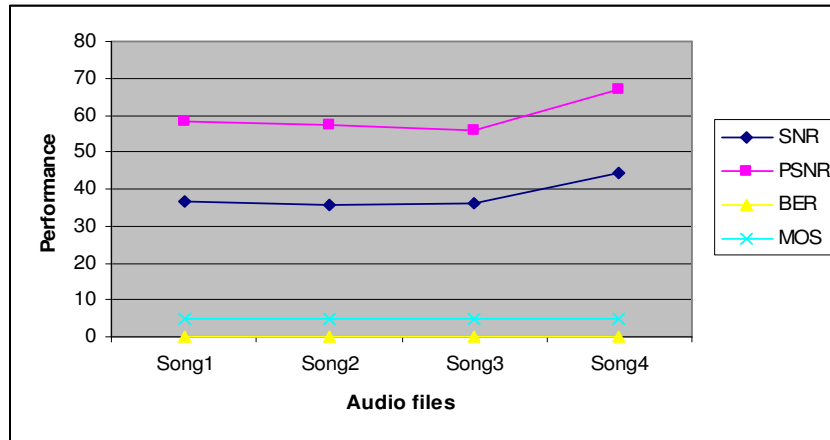| Rating	Impairment	Quality
5	Imperceptible	Excellent
4	Perceptible, not annoying	Good
3	Slightly annoying	Fair
2	Annoying	Poor
1	Very annoying	Bad

Figure 4. Performance for different audio signals

The quality rating (Mean opinion score) is computed by using equation (2) and BER is represented by equation (3).

$$Quality = \frac{5}{1 + N * SNR} \tag{2}$$

Whereas N is a normalization constant and SNR is the measured signal to noise ratio.

$$BER = \frac{100}{l} \sum_{n=0}^{l-1} \begin{cases} 1, & W'n=Wn \\ 0, & W'n \neq Wn \end{cases} \tag{3}$$

4. CONCLUSION AND FUTURE WORK

In this paper, perceptual encoding has been applied to systematic representing the fraction part of sampled values of song signal or a portion of song signal as well as embedding authenticated code with the lower bit pattern of magnitude values in selected frequency region has been proposed which will not affect the song quality but it will ensure to detect the distortion of song signal characteristics. Additionally, the proposed algorithm is also very easy to implement.

This technique is developed based on the observation of characteristics of different songs with human audible capability and an approach is also made for estimating the embedded extra data limit with the help of Shannon's limit in the channel encoding scheme. It also can be extended to embed an image or audio into an audio signal instead of text and discrete sampled values. The perfect estimation of percentage of threshold numbers of sample data of song that can be allow to change for a normal conditions will be done in future with all possibilities of errors in song signal processing.

ACKNOWLEDGEMENTS

REFERENCES

[1] Mondal, Uttam Kr., Mandal, J.K. (2010) A Practical Approach of Embedding Secret Key to Authenticate Tagore Songs(ESKATS), Wireless Information Networks & Business Information System Proceedings (WINBIS'10), ISSN 2091-0266,organized by Rural Nepal Technical Academy (Pvt.) Ltd , Nepal, Volume 6, Number 1, pp 67-74.

[2] Mondal, Uttam Kr., Mandal, J.K. (2010) A Novel Technique to Protect Piracy of Quality Songs through Amplitude Manipulation (PPAM), International Symposium on Electronic System Design (ISED 2010), ISBN 978-0-7695-4294-2 ,pp 246-250.

[3] Xiaoxiao, Dong, Mark, F., Bocko, Zeljko Ignjatovic(2004) Data Hiding Via Phase Manipulation of Audio Signals , IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004) , ISBN 0-7803-8484-9 ,Vol 5, pp 377-380.

[4] Erten, G., Salam, F. (1998) Voice Output Extraction by Signal Separation", ISCAS '98 ,ISBN 07803-4455-3,Vol 3, pp 5 - 8.

[5] Katzenbeisser, S., Petitcolas F.A.P. (2000) Information Hiding Techniques for Steganography and Digital Watermarking, Artech House, Norwood, MA, ISBN 978-1-58053-035-4.

[6] Arnold, M.(2000) Audio watermarking: Features, applications and algorithms, IEEE International Conference on Multimedia and Expo ,New York, NY, Vol 2, pp. 1013-1016.

[7] Uttam Kr. Mondal, J K. Mandal (2012) Secret Data Hiding within Tolerance Level of Embedding in Quality Songs (DHTL), Second International Conference on Computer Science, Engineering and Application (ICCSEA 2012), Vol 2, ISSN 1867-5662, ISBN 978-3-642-30110-0, pp 753-761.

[8] http://en.wikipedia.org/wiki/Psychoacoustic last accessed at 9:20 PM on 06/09/2012.

[9] Stan Z. Li : Content-Based Audio Classification and Retrieval Using the Nearest Feature Line Method, IEEE Transactions on Speech and Audio Processing, Vol 8, No 5, September 2000, pp 619-625.

[10] Uttam Kr. Mondal, J K. Mandal: An Analytical Approach to Generate Unique Song Signal (AUSS), Third International Conference on Computer Science & Information Technology(CCSIT-2013) vis Lecture Notes of the Institute for Computer Science, Social Informatics and Telecommunications Engineering(LNICST), Springer, ISSN No.1867-8211,Feb.18-20, 2013, Bangalore, India.

Authors

Uttam Kr. Mondal, has received his Bachelor of Engineering (B.E) degree in Information Technology in 2004 and Master of Technology (M.Tech) in Information Technology in 2006 from University of Calcutta, India. He has now working as an Asst. Professor in department of Computer Science & Engineering and Information Technology in College of Engg. & Management, Kalaghat, West Bengal, India. His research areas include cryptography & Network Security, Audio signal authentication. He has 26 publications in National and International conference proceedings and journal.

Jyotsna Kumar Mandal, M. Tech.(Computer Science, University of Calcutta),Ph.D.(Engg., Jadavpur University) in the field of Data Compression and Error Correction Techniques, Professor in Computer Science and Engineering, University of Kalyani, India. Life Member of Computer Society of India since 1992 and life member of cryptology Research Society of India. Dean Faculty of Engineering, Teachnology & Management, working in the field of Network Security, Steganography, Remote Sensing & GIS Application, Image Processing. 26 years of teaching and research experiences. Nine Scholars awarded Ph.D. and 8 are pursuing. Total number of publications 277 in addition to five books published by LAP Germany.