AN ANALYTICAL APPROACH TO GENERATE UNIQUE SONG SIGNAL (AUSS)

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

Embedding uniqueness in characteristics of song signal and accustoming changes of environment is one of the challenging issues for researchers with maintaining its audible quality. Researchers are modifying or manipulating audio signal properties for generating uniqueness in content such a manner that will not vary so much in changed environment or changes can be easily defined due to unique structure of song signal. In this paper, an approach has been made based on defining a symmetric structure of song signal, followed by some secret code embedding in a specified manner will not alter the trade off ratio of embedding/modifying data but provide uniqueness in properties, even retain the properties in changing environment/ format. Therefore, authentication of song signal is easily achieved with these self manipulated properties. 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

Audio authentication is one of the primary requirements of music industries. Alternating the content of originally generated song/music with the use of mixing or editing technology, new versions are released; original investors are falling to generate projected revenue. A prolong demands to produce an self authenticated song signal which can be protected from any kind of malicious attack, even if any part of its content alter directly affect its audible quality. i.e., an integrated continuous authenticated song signal is required which itself carries a hormonal relationship among all its basic properties [1, 2]. Alternation of any of them directly affect over its audible quality. Therefore, a quality estimation is also required such that modified self authenticate song signal will not alter as much as tolerance level of audible quality of song signal. The modified content of song signal will contain with marginal values of all basic properties,

Jan Zizka (Eds) : CCSIT, SIPP, AISC, PDCTA - 2013
pp. 209–216, 2013. © CS & IT-CSCP 2013 DOI : 10.5121/csit.2013.3623

changing any part of any property will create alternation in audible quality [4]. The effect of underlying relationship will hold even after conversion in another format of modified song signal after embedding authenticated code.

In this paper, a framework for protecting originality of a particular song with reconstructing its content with the help of amplitude coding of sampled values followed by passing a secret code in bit level of its content. The reformation in content values as well as authenticating code will use to detect and identify the original song from similar available songs. It is experimentally observed that added extra values will not affect the song quality but provide a level of security to protect the piracy of song signal.

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

2. THE TECHNIQUE

The scheme fabricates the secret key with help of perceptual coding technique in the fraction part of sample values of song signal followed by encoding bit level value for generating unique signature. 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 AUSS-ESV and AUSS-SCB are proposed as double security measure, the details of which are given in section 2.1 and section 2.2 respectively.

2.1 Lossy Encoding of Sampled Values (AUSS – 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. The encoding technique is rounding off the value of fraction part of sampled values considering its near ceiling boundary values (determine 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 Constructing Secret Code at Bit-level (AUSS – SCB)

Representation of magnitude values in bit level is to be justified, estimation has to be made for the specific position of bit representation, having zero value for binary representation of decimal part of magnitude value. Send a secret message with all positions containing "1" value for the particular position, if more than one region found, choose most appropriate region as comparing the content of magnitude values in bit stream. Then, a self authenticated message will be embedded without violating its audible quality of song signal. The embedding process is as follows.

i. Apply FFT to find frequency component of song signal.
ii. Represent the decimal part of magnitude values into binary stream for both channels of stereo type song.
iii. Scan all binary streams and find the bit level positions form LSB to MSB to find most 1's positions for magnitude values.
iv. Compare two channels most "1" values positions and select a position in which two channels converted binary values for that position are approximately equal. Then make them exactly equal for both channels. It will be best suited if we able to find the position in the nearby region of middle position of LSB and MSB for less attack on this secret code under application of any compression technique of the original song.
v. Another bit position may be considered at nearby MSB for tracking the alter bit positions of step (iii) which one should be most "0" value position for all sampled values but only hold "1" at alter positions. Additionally another bit position may be considered for selecting particular channel in similar way.
vi. Apply inverse FFT to get back the sampled values of modified song signal.

In case of mono type song, step (iii) is not required, but for the both cases error estimation must be applied for getting appropriate trade-off ratio [7].

2.3 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 the 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 2.2. 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].

2.4 Authentication

Estimated ceiling value of fraction part of sampled values and symmetric representation in bit level for a particular region for both channels of stereo type song creates a secure code that will use to identify the original song. The information at selected k^{th} bit level position of two channels carries unique signature for original song signal. The number of security strings of 1's may be increased as required or based upon the acceptance trade-off ratio of original song and impurities (alternated bit values) [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.

2.5 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 position of higher bit level.

Step 3: Apply step 2 for the all positions of the selected region 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 AUSS 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. AUSS is applied on this signal and as a first step of the procedure which is performed AUSS 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.

Table 1. 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
...	...	...

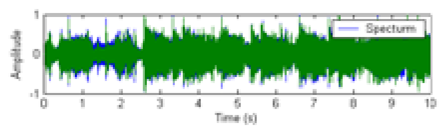


Fig. 1. Original song ('100 Miles From Memphis', sang by

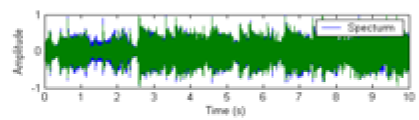


Fig. 2. Modified song with secure code

3.1.2 Modified song after bit level encoding and rounding off the fraction part of magnitude values (10 seconds)

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

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.

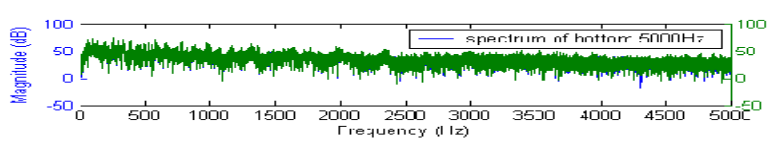


Fig. 3. The difference of magnitude values between signals shown in figure 1 and 2.

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)

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

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 $P = E[X_k^2]$ [16-bit stereo audio signals sampled at 44.1 kHz],

Where, 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 (0.0131), will not affect over all song audible quality. Because, only about 1030 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] 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. 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	Value using DHPMA [3]
1	MD	0.0154	3.6621e-004
2	AD	0.0102	2.0886e-005
3	NAD	0.0101	0.0063
4	MSE	5.3738e-006	1.4671e-009
5	NMSE	2.2367e+004	8.4137e-005

Table 3. SNR and PSNR

Sl No	Statistical parameters for Differential distorstion	Value using AUSS	Value using DHPMA [3]
1	Signal to Noise Ratio (SNR)	38.0041	40.7501
2	Peak Signal to Noise Ratio (PSNR)	54.407	45.4226

Table 4. Representing NC and QC

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

Table 5. Showing SNR, PSNR BER, MOS

Audio (10s)	SNR	PSNR	BER	MOS
Song1	41.2265	65.2981	0	5
Song2	38.0041	54.407	0	5
Song3	37.0018	50.6411	0	5
Song4	43.886	64.9013	0	5

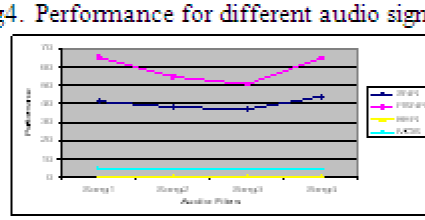
Where N is a normalization constant and SNR is the measured signal to noise ratio.

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

Fig4. Performance for different audio signals



The quality rating (Mean opinion score) is computed by using equation (2).

$$Quality = \frac{5}{1 + N * SNR} \qquad (2)$$

4. CONCLUSION AND FUTURE WORK

In this paper, an algorithm for encoding the fraction part of sampled values with perceptual coding technique as well as embedding some secret code in bit level operation 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 characteristics 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 into an audio signal instead of text and audio. 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.

REFERENCES

[1] Mondal, Uttam Kr., Mandal, J.K.: 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(2010).

[2] Mondal, Uttam Kr., Mandal, J.K. : 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(2010).

[3] Xiaoxiao, Dong, Mark, F., Bocko, Zeljko Ignjatovic: 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(2004).

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

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

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

[7] Uttam Kr. Mondal, J K. Mandal: 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.