Journal of Engineering Science and Technology
Vol. 12, No. 6 (2017) 1627 - 1639
© School of Engineering, Taylor's University

SECUREROBUST AND HYBRID WATERMARKING FOR SPEECH SIGNAL USING DISCRETE WAVELETTRANSFORM DISCRETE COSINE TRANSFORM ANDSINGULAR VALUE DECOMPOSITION

AMBIKA DORAISAMY*, RADHA VENKATACHALAM

Department of Commerce, Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore - 641043, Tamil Nadu, India
*Corresponding Author: ambika.avinuty16@gmail.com

Abstract

A digital watermark is defined as inaudible data, permanently embedded in a speech file for authenticating the secret data. The main goal of this paper is to embed a watermark in the speech signal without any degradation. Here the hybrid watermarking is performed based on the three techniques such as Discrete Cosine Transform (DCT) with Singular Value Decomposition (SVD) and Discrete Wavelet Transform (DWT) and it is optimized by performing the separation of speech and silent regions using a voice activity detection algorithm. The performances were evaluated based on Peak Signal to Noise Ratio (PSNR) and Normalized Cross Correlation (NCC). The result shows that the optimization method performs better than the existing algorithm and it is robust against different kinds of attacks. It also shows that the algorithm is efficient in terms of robustness, security, and imperceptibility and also the watermarked signal is perceptually similar to the original audio signal.

Keywords: Speech signal, Watermarking, Robustness, Security, Imperceptibility.

1. Introduction

Audio watermarks are special signals embedded into digital audio. These schemes rely on imperfections of the human auditory system. Research in audio watermarking is not as mature, compared to research in image and video watermarking [1]. This is due to the fact that, the human auditory system is much more sensitive than the human visual system, and that inaudibility is much more difficult to achieve than invisibility for images. Furthermore, audio signals are represented by much less samples per time interval, which indicates that the amount of information capacity that can be embedded robustly and inaudibly in audio files is much lower than the amount of information that can be embedded in

Nomenclatures

$M*N$	Size of the watermark
M	Mean Value
P	Percentage of energy of each coefficient
q_i	Quantization
$W'(i,j)$	Pixel of extracted watermark
X	Maximum absolute square value of the signal
$\|\|x-r\|\|^2$	Energy of the difference between original & reconstructed signal

Greek Symbols

σ_i	Standard deviation
ψ	Mapping Function

Abbreviations

AV	Amplitude Variation
CAR	Confidentiality Availability and Reliability
DCT	Discrete Cosine Transform
DCWS	DWT and SVD
DDSVRS	DWT+SVD+DCT with VAD and Embedding Region
DDSWA	DWT+SVD+DCT Watermarking Algorithm
DS	Down-Sampling
DSRS	DWT+SVD Watermarking Algorithm with Embedding Region
DSVAD	DWT+SVD Watermarking Algorithm with VAD
DSWA	DWT+SVD Watermarking Algorithm
DWT	Discrete Wavelet Transform
DWS	DWT + SVD based Watermarking
EA	Echo Addition
EDCWS	Enhanced DCT + DWT + SVD based Watermarking
GN	Gaussian Noise
IDCT	Inverse Discrete Cosine Transform
LPF	Low-Pass Filtering
MC	MP3 Compression
MFCC	Mel frequency cepstrum coefficient
NA	No attacks
NCC	Normalized cross correlation
PSNR	Peak Signal to Noise Ratio
RA	Reverse Amplitude
RQ	Re-Quantization
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SNR	Signal to Noise Ratio
US	Up-Sampling
W1&W2	Copyright 1 and Copyright 2
VAD	Voice Activity Detection

visual files. Nonetheless, many audio watermarking techniques have been proposed in literature in recent years. Most of these algorithms attempt to satisfy watermarking requirements by exploiting the imperfections of the human auditory system. They exploit the fact that human auditory system in insensitive to small amplitude changes, either in the time-domain or frequency-domain, to embed watermark information.

Moving towards secure environment, several business and industry applications require secure standards that are not normally provided during transmission. This work can be applied in applications such as Governments, military, corporations, financial institutions, hospitals and private businesses exchange confidential information frequently. Most of this information is collected, processed and stored on computers and transmitted across networks to other computers. If this information falls into wrong hands, it could lead to lost business, law suits, identity theft or even bankruptcy of the business. With cyber-crime on the rise, protecting information and assets belonging to both corporate and government organizations is vital. Protecting confidential information during communication is a mandatory requirement and in many cases also an ethical and legal requirement. The main goal of this research is to secure a speech data through the effort of watermarking without degrading the speech quality. Here the hybrid techniques are used because as we move towards the secure environment, several business and industry applications require secure standards that are not normally provided during transmission.

Digital speech watermarking is a robust way to hide and thus secure data like audio from any intentional or unintentional manipulation through transmission. The additional information which is embedded in the host signal should be extractable, and must resist various intentional and unintentional attacks. Nowadays there is a great demand for multimedia applications which has paved the way for secure communication. For digital audio, it is difficult to secure and protect an author’s work from being copied. In wireless channel, it is necessary to provide protection, to avoid unauthorized modifications and disclosure when transmitting any type of data such as speech and text. The hiding of information is unlike cryptography and can be used in various applications such as copyright protection, broad cast monitoring and covert communication.

Digital speech watermarking is used to hide and protects data like audio from being manipulated through transmission. The additional information which is embedded in the host signal should be extractable and must resist various intentional and unintentional attacks. Many watermarking schemes are proposed [2, 3] and each aiming to develop robust watermarks that protects digital contents during transmission. However, the continuing revolution in the communication medium is demanding and as a result, it has become imperative to improve watermarking techniques that can satisfy the property of Confidentiality, Availability and Reliability (CAR) along with maximum transparency, capacity and robustness. Search for techniques to satisfy the above desired properties is still an active research area and is the focus of the present research work.

The main goal of this work is to find a technique that can simultaneously protect, preserve security without destroying or modifying the content of the digital speech signal. In the existing system, the authors Makbol and Khoo [4] proposed a scheme using DWT and Singular Value Decomposition for watermarking image data. The efficiency of this model reduced when applied to speech data and there was audible distortion in the water marked signal if speech has silent region. Generally, it is not robust when used with speech data. In this research work, the Makbol and Khoo model is improved to solve the above-mentioned problems. The paper is organized as follows; in section 2 the Proposed Methodology is discussed, Section 3 simplifies the Experimental Results and Discussion, finally the Conclusion is summarized in Section 4.

2. Proposed Method

The proposed methodology was modified to include the following optimization operations:

- Separation of speech and silent region
- Selection of embedding region in speech region
- Replace DWT + SVD with DWT + DCT + SVD algorithm

The separation of speech and silent regions was performed using a voice activity detection algorithm. This algorithm begins by analysing the input speech signal in the form of fixed size window by extracting MFCC. Then, a neural network based algorithm is used to detect the speech and non-speech signals both in noisy and clean segment of the input speech signal. The Signal-to-Noise Ratio (SNR) is used for this purpose. When an interchange from speech to non-speech (or vice-versa) occurs, the energy change in Wavelet Coefficients is used to determine the start and end points of the speech signals. The obtained speech signal (without silence segment) is then normalized. The algorithm then uses a hybrid watermarking algorithm that combines DWT, DCT (Discrete Cosine Transformation) and SVD to watermark the cover speech signal.

The proposed watermarking algorithm enhances the operation of an existing Discrete Wavelet Transformation (DWT) based algorithm that combines DWT and Singular Value Decomposition. This algorithm is termed as DWS (DWT + SVD based Watermarking) algorithm in this research work. The main goal of this work is to enhance DWS algorithm, by replacing DWT + SVD combination with DCT (Discrete Cosine Transformation), DWT and SVD (DCWS) and speech activity based watermarking. The proposed algorithm is termed as EDCWS (Enhanced DCT + DWT + SVD based Watermarking) algorithm in this research work and it consists of the following steps.

Step 1: Voice Activity Detection
Step 2: Application of 2-level DWT on Cover Data
Step 3: Application of DCT on Step 2 Results and Watermark Data
Step 4: Application of SVD on Results of Step 3
Step 5: Embed Watermark in Cover
Step 6: Perform Inverse Transformations to obtain watermarked data

2.1. Discrete cosine transform

Discrete cosine transform (DCT) translates the information from spatial domain to frequency domain to be represented in a more compact form. Its stochastic properties are similar to Fourier transform and consider the speech signal to be a time invariant or stationary signal. The DCT is a special case of Discrete Fourier Transform (DFT) in which the sine components have been eliminated leaving only the cosine terms [5]. Both DCT and IDCT are orthogonal, separable and real transforms. Separable means that the multidimensional transform can be decomposed into successive application of one-dimensional transforms in the appropriate directions. Similarly orthogonal means if the matrices of DCT and

IDCT are non-singular and real then their inverse is obtained merely by applying transpose operation. Like Fourier transform, DCT also considers the input sampled data to be a time invariant or stationary signal. The 8-point 2-D DCT and IDCT to generate 8x8 data matrices are calculated as:

$$X_{k,l} = \frac{c(k)c(l)}{4} \sum_{m=0}^{7}\sum_{n=0}^{7} x_{m,n} \cos\left(\frac{(2m+1)k\pi}{16}\right)\cos\left(\frac{(2n+1)l\pi}{16}\right) \quad (1)$$

where k, $l = 0, 1, \ldots, 7$.

$$x_{m,n} = \sum_{k=0}^{7}\sum_{l=0}^{7} \frac{c(k)c(l)}{4} X_{k,l} \cos\left(\frac{(2m+1)k\pi}{16}\right)\cos\left(\frac{(2n+1)l\pi}{16}\right) \quad (2)$$

where m, $n = 0, 1, \ldots, 7$ and $c(k), c(l) = \begin{cases} 1/\sqrt{2}, k \,\&\, l = 0 \\ 1, otherwise \end{cases}$ (3)

2.2. Discrete wavelet transform

The WT has a good time and poor frequency resolution at high frequencies, and good frequency and poor time resolution at low frequencies. Discrete wavelet transform (DWT), which transforms a discrete time signal to a discrete wavelet representation, here the first step is to discrete the wavelet parameters, which reduce the continuous basis set of wavelets to a discrete and orthogonal/orthonormal set of basis wavelets. The 1-D DWT is given as the inner product of the signal *x(t)* being transformed with each of the discrete basis functions.

$$W_{m,n} =< x(t), \psi_{m,n}(t) >; m, n \in Z \quad (4)$$

The 1-D inverse DWT is given as:

$$x(t) = \sum_{m}\sum_{n} W_{m,n}\psi_{m,n}(t); m, n \in Z \quad (5)$$

The 1-D DWT can be extended to 2-D transform using separable wavelet filters. With separable filters, applying a 1-D transform to all the rows of the input and then repeating on all of the columns can compute the 2-D transform.

2.3. Singular value decomposition

The singular value decomposition [6, 7] is a more general method that factors any m x n matrix A of rank r into a product of three matrices, such that

$$A = U\sum V^T \quad (6)$$

where U is m x m (left singular vectors), $\sum$ is m x n and V is n x n (right singular vectors). U and V are both orthonormal matrices (i.e., $UU^T = I_m$ and $VV^T = I_n$) and $\sum$ is an m x n matrix where the non-negative and descending order elements along the diagonal of the left/top quadratic sub block are the singular value $\sigma_1 \geq \sigma_2 \geq \sigma_3 \ldots \geq \sigma_r \geq 0$ of A. All other elements of $\sum$ are 0.If $r<\min(m, n)$, that is, the matrix A is not full-rank, then only r singular values are greater than 0. A full rank decomposition of A is usually denoted as $A_r = U_r$ *Erode* V_r^T. The singular values are always real numbers. If the matrix A is real, then U and V are also real. The reduced rank SVD to A can be found by setting all but the first k largest singular values equal to zero and using only the first k columns of U and V. This is denoted using Eq. (7).

$$A_k = U_k \sum_k V_k^T \quad (7)$$

2.4. Voice activity detection

An important step of speech communication is Voice Activity Detection (VAD). In this research work, a VAD system is designed to identify areas which can be used to insert watermark without degrading the quality of the speech signal. The usefulness of VAD has already been well-proven in the field of speech recognition [8-10], speech compression [11], speech enhancement [12] and in general, all telephone applications [13].The proposed VAD consists of the steps presented in Fig. 1. The audio input signal is processed within a fixed length of non-overlapped window. The MFCC features are extracted, and two SVM classifiers are trained to speech/non-speech signals.

The first classifier is used to detect speech under noisy condition, while the second is used to detect speech in noise-free speech signals. During training, white noise was added to the speech signals with SNR range from 30dB to 40dB because SNR directly impacts the performance of a wireless LAN connection. A higher SNR value means that the signal strength is stronger in relation to the noise levels, which allows higher data rates and fewer retransmissions – all of which offers better throughput. A lower SNR requires wireless LAN devices to operate at lower data rates, which decreases throughput. A SNR of 30 dB, for example, may allow an 802.11g client radio and access point to communicate at 24 MBPS. In the next step, the SNR of the current windowed signal to the fixed cleaned windowed non-speech signal is estimated.

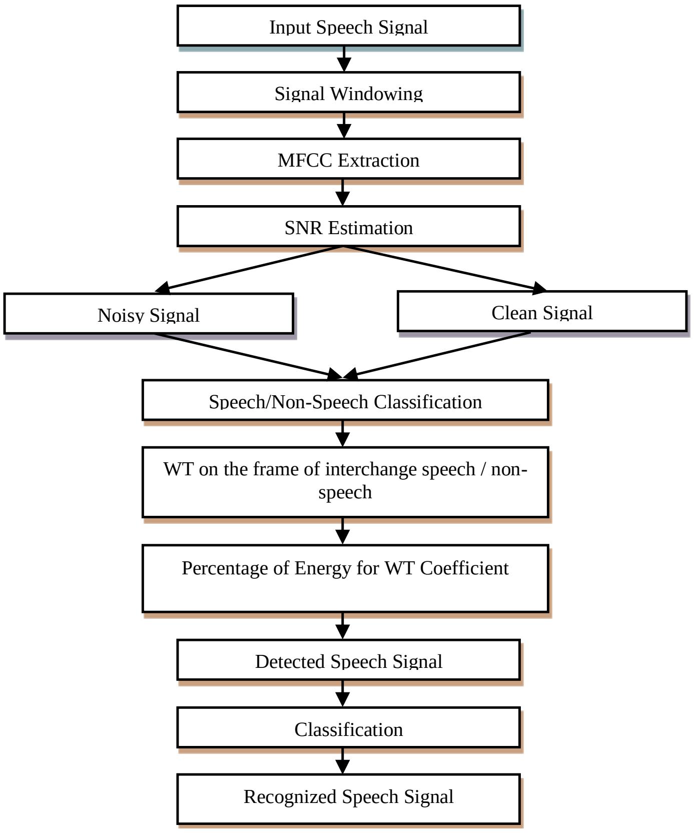


Fig. 1. VAD algorithm.

For this purpose, a SNR threshold is used to decide on noisy and clean data. This value was set to 0.05 after conducting repeated experiments with different threshold values. After estimating clean and noisy data, the features extracted are given to the appropriate trained classifier. The output of 0 (zero) indicates the presence of speech and 1 (one) indicates non-speech signals. When a change from one to zero occurs (a change from non-speech to speech) or vice versa occurs, the interchange windows which is within the boxes labelled is sent for further signal starting/ending point localization. For this, the interchange frame is wavelet decomposed to obtain its coefficients.

The percentage of energy of each coefficient is then computed using Eqs. (8) and (9).

$$P = abs(coeff \times coeff) \tag{8}$$

$$Energy\ percentage = \frac{p}{sumP} \times 100 \tag{9}$$

Using these values, a scalogram is constructed to represent the energy changes of the interchange for signal starting and ending point. From the scalogram, the position of maximum energy is determined. Subsequently, according to the energy which gradually reduces as the contour spreading from the maximum point, the coordination where the energy reaches the predefined threshold (3.3%) located at the same horizontal position as the maximum energy is noted as the starting/ending point speech signal. The starting and ending points are plotted on the respective scalogram. Finally, the resultant of the speech signal localization is obtained.

2.5. Selection of embedding region

This procedure also uses the high-energy regions for embedding the watermark signals. During experimentation, it was found that short time frames around all the strong local energy peaks are places which produce minimum distortion and hence are used for embedding the watermark. For this purpose, the signal is first smoothed and then all the strong peaks are estimated by calculating gradient and slope. The frames around these peaks are selected to insert the watermark.

2.6. EDCSW algorithm

The EDCSW algorithm embeds watermark into cover speech data into three stages Fig. 2.

(i) Prepare cover speech data
(ii) Prepare watermark image
(iii) Embed watermark into cover data

The cover speech data is first divided into frames of size 8K. The performance of VAD as explained above. As embedding watermark into non-speech signal will not degrade speech quality, this area is selected for watermark embedding. Then a two-level DWT is applied, which gives four subbands, namely, LL, LH, HL and HH. The HH subband is again decomposed using DCT. Map the DCT coefficients into four quadrants by zigzag scanning. The SVD algorithm is then applied to each quadrant. Let the result be denoted as SVD$_{c1}$, SVD$_{c2}$, SVD$_{c3}$ and SVD$_{c4}$. The same procedure is used to obtain the SVD values of the watermark image. Let the result be denoted as SVD$_{w1}$, SVD$_{w2}$, and SVD$_{w3}$ and SVD$_{w4}$. Replace SVD$_{ci}$ coefficients with SVD$_{wi}$ coefficients using Eq. (10). Let these be denoted as SVD$_{\mathrm{Newqi}}$ (i = 1, 2, 3, 4 representing the quadrants).

$$SVD_{Newq_i} = SVD_{Ci} + SVD_{Wi} \tag{10}$$

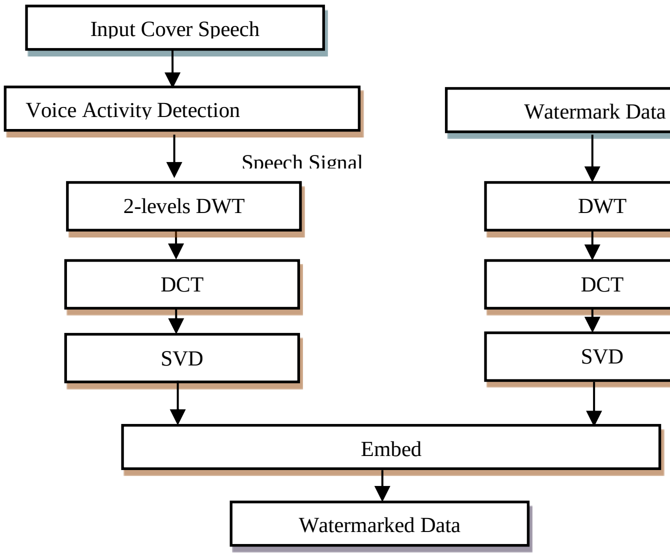


Fig. 2. Steps in EDCSW algorithm.

The coefficients are mapped from zigzag scanning to original position matrix. Application of Inverse DCT (IDCT) to this matrix, followed by Inverse DWT, produces the watermarked signal. Before embedding, a quantization for each quadrant is estimated and is stored as secret key. The quantization q_i for i^{th} quadrant is computed using Eq.(11).

$$q_i = q_m + (q_M - q_m)\frac{S_i - S_{\min}}{S_{\max} - S_{\min}} \tag{11}$$

where $S_i = m_i \ x \ (\sigma_i)^{1.25}$, m and σ_i are the mean value and standard deviation of i^{th} quadrant. Parameters q_m and q_M represent the minimum and maximum quantization step values respectively. The parameter q_i denotes the quantization step of quadrant i and is proportional to the numerical value S_i. These quantization parameters are saved as secret keys for watermark recovery. Figure 3 presents the watermark extraction procedure, while applies DWT first to obtain four subbands, LL, LH, HL and HH. DCT is then applied to HH subbands to obtain DCT coefficients. These coefficients are mapped to four quadrants by zigzag scanning. Apply SVD to obtain SVD$_{w1}$, SVD$_{w2}$, SVD$_{w3}$ and SVD$_{w4}$. Extract watermarked bits from these SVDs using Eq.(12).

$$SVD_{wi}' = (SVD_{Newq} - SVD_{Ci}) \tag{12}$$

The next step of the algorithm, reconstructs SVD matrix for each quadrant. This is followed by the application of IDCT and IDWT to obtain the watermark removed cover data.

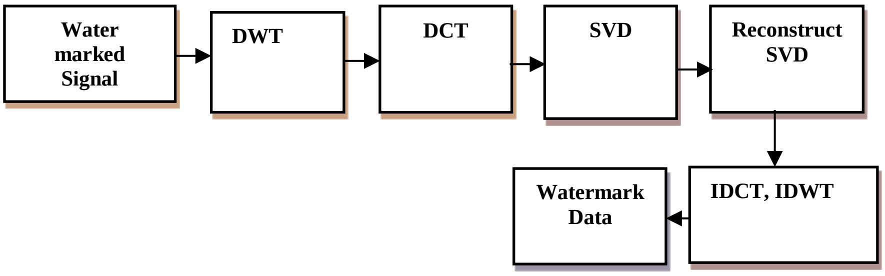


Fig. 3. Watermark extraction procedure of EDCSW.

3. Experimental Results of Watermarking Algorithm

The performance was compared using standard parameters, namely, capacity, perceptibility, robustness, security and complexity. The **capacity** is defined as the number of bits of watermark image is embedded into the speech sample. As the size of the watermark is fixed (32 x 32 pixels), this parameter is taken as efficient, if it works well for all the other watermark characteristics. The **perceptibility** of a watermarking algorithm is used to calculate the perceptual distortion between the original and watermarked speech signal. A digital watermark is called imperceptible if the original cover signal and the watermarked signal are perceptually indistinguishable. A digital watermark is called perceptible if its presence in the watermarked signal is noticeable. Both capacity and perceptibility is analysed using the Peak Signal to Noise Ratio (PSNR).

The **robustness and security** of the proposed watermarking technique is analysed using a set of operations and manipulations made to the watermarked speech signal. The manipulations made are termed as 'Watermark Attack' and is defined as an attempted modification or damage of digital data to identify the hidden watermark. The attacks used in the study are Down-Sampling, Up-Sampling, Gaussian Noise, Low-Pass Filtering, Re-Quantization, MP3-Compression, Echo Addition, Reverse Amplitude, and Amplitude Variation. The parameter used to analyse the proposed algorithm with respect to attacks is Normalized correlation (NCC) and is calculated using Eq. (13).

3.1. Normalized cross correlation (NCC)

NCC gives the similarity measure between the original and extracted watermark. The maximum value of NCC is 1.

$$NCC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} W(i,j)W'(i,j)}{\sum_{i=1}^{M}\sum_{j=1}^{N} W(i,j)^2} \tag{13}$$

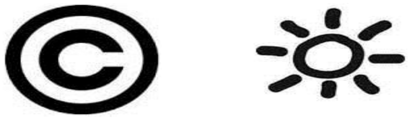
where *W(i, j)* denotes the $(i, j)^{th}$ pixel of original watermark, *W'(i, j)* is the $(i, j)^{th}$ pixel of extracted watermark. *M* x *N* is the size of watermark.

3.2. Peak Signal to Noise Ratio (PSNR)

PSNR is often used as a quality measurement between the original and the decompressed signal. The higher PSNR is the better quality of the decompressed signal. It is estimated according to Eq. (14).

$$PSNR = 10\log_{10}\frac{NX^2}{\left\|x - r^2\right\|} \tag{14}$$

where N is the length of the reconstructed signal, X is the maximum absolute square value of the signal x and $\| x\text{-}r\|^2$ is the energy of the difference between the original and reconstructed signals.



(a) W1 **(b) W2**

Fig. 4. Selected watermark images.

The experiments were conducted using 20 watermark images (32 x 32 pixels), stored in JPEG format. Five speech signals from five speakers denoted as Speaker 1, Speaker 2 and ...Speaker 5 and two watermark images namely (W1 andW2) were selected randomly and used during discussion. The two selected watermark images are shown in Fig. 4. The experiments first analyse the efficiency of the proposed algorithm along with its optimization procedure. Next, the robustness of the algorithm on various attacks are analysed and determined. The various experimental results as given in Fig. 5 for PSNR and Fig. 6 for NCC compared the performance of the existing watermarking algorithm with the enhanced algorithm. Each optimization operation, treated as on/off option, was incorporated separately to the existing scheme to analyse its effectiveness on watermarking performance. To analyse the robustness of the enhanced watermarking algorithm, the second set of experiments analyses its efficiency in the presence of the selected attacks. Figure 7 presents the NCC for the watermarking1 and Fig. 8 presents the NCC for the watermarking2 for the enhanced method DDSVRS. Similarly, Fig. 9 presents the PSNR for the watermarking1 and Fig. 10 presents the PSNR for the watermarking2 for the enhanced method DDSVRS. Here the watermarked signals are attacked with the selected nine attacks. From this set of experiments, the following facts were ascertained.

- **Robustness Characteristic**: From the results pertaining to attacks, it can be seen that the enhanced algorithm is robust against all the considered attacks in terms of NCC and BER.
- **Security:** The enhanced algorithm provides security in a multiple step fashion. First, the watermark is embedded only in the presence of region where voice activity is there. Finally, the usage of SVD is very robust and small modifications do not affect the quality of the speech data.
- **Imperceptibility Characteristic:** The reported results with respect to NCC and BER show that the enhanced DDSVRS watermarking algorithm does not degrade the signal even in the presence of attacks. The results

also show that the algorithm has improved the performance of watermarking algorithm when compared to the existing watermarking algorithm. This proves that the DDSVRS watermarking algorithm has improved imperceptibility characteristic.

The result shows that the enhancement operations, voice activity detection, region selection and combined DWT, DCT and SVD transformations, have improved the process of watermarking speech data in a positive manner. The results also show that the proposed algorithm is efficient in terms of robustness, security, imperceptibility and speed. The performance trend reported was the same for all the images and speech samples of all datasets.

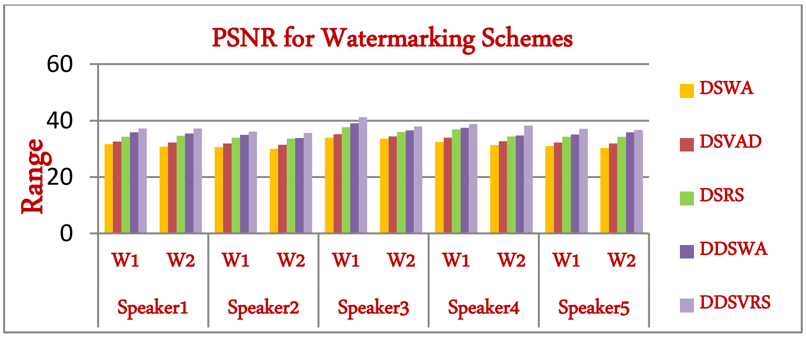


Fig. 5. PSNR for watermarking schemes.

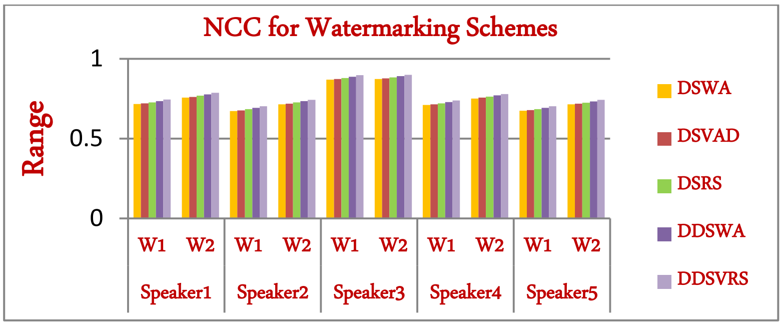


Fig. 6. NCC for watermarking schemes.

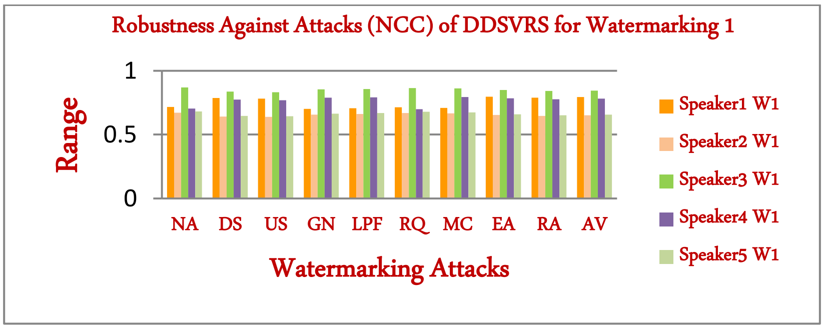


Fig. 7. NCC of DDSVRS for watermarking1.

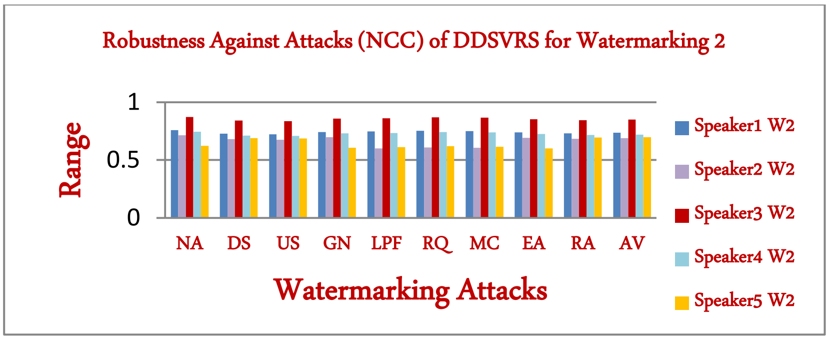


Fig. 8. NCC of DDSVRS for watermarking2.

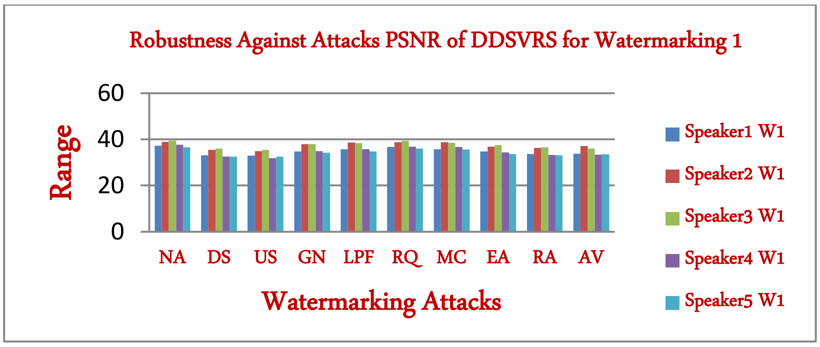


Fig. 9. PSNR of DDSVRS for watermarking 1.

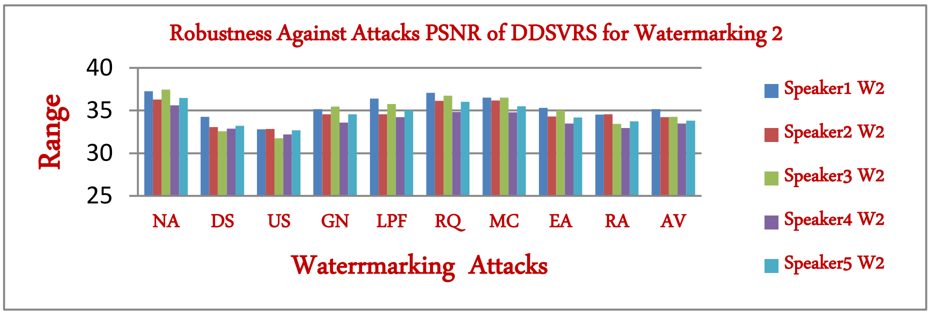


Fig. 10. PSNR of DDSVRS for watermarking 2.

4. Conclusion

The algorithm is optimized by performing the separation of speech and silent regions using a voice activity detection algorithm. To analyse the robustness of watermarking, nine different attacks were used. They are Down-Sampling, Up-Sampling, Gaussian Noise, Low-pass Filtering, Re-quantization, MP3 Compression, Echo Addition, Reverse Amplitude and Amplitude Variation. Upon experimentations, it was found that all the enhanced techniques performed satisfactorily and was able to protect the speech signal in an efficient manner with little degradation to speech signal quality. And all so the hidden watermark data is robust against different kinds of attacks. The watermarked signal is perceptually similar to the original audio signal and produced high quality output.

References

1. Arnold, M. (2003). Attacks on digital audio watermarks and countermeasures. *Proceedings of Third IEEE International Conference on WEB Delivering of Music*, 1-8.
2. Chen, N.; and Zhu, J. (2007). Robust speech watermarking algorithm. *Electronics Letters*, 43(24), 1393-1395.
3. Nematollahi, M.A.; Al-Haddad, S.A.R.; Doraisamy, S.; and Saripan, M.I.B. (2012). Digital audio and speech watermarking based on the multiple discrete wavelets transform and singular value decomposition. *Sixth Asia Modelling Symposium*, 109-114.
4. Makbol, N.M.; and Khoo, B.E. (2013). Robust blind image watermarking scheme based on redundant discrete wavelet transform and singular value decomposition. *AEU - International Journal of Electronics and Communications*, 67(2), 102-112.
5. Mandyam, G.; Ahmed, N.; and Magotra, N. (1997). Lossless image compression using the discrete cosine transform. *Journal of Visual Communication and Image Representation*, 8(1), 21-26.
6. Demmel, J.W. (1997) Applied numerical linear algebra. SIAM
7. Berry, M.W.; Drmac, Z.; and Jessup, E.R. (1999). Matrices, vector spaces and information retrieval. *SIAM Review*, 41(2), 335-362.
8. Ramirez, J.; Segura, J.C.; Gorriz, J.M.; and Garcia, L. (2007). Improved voice activity detection using contextual multiple hypothesis testing for robust speech recognition. *IEEE Transactions on Audio, Speech and Language Processing*, 15(8), 2177-2189.
9. Gorriz, J.M.; Ramírez, J.; Lang, E.W.; Puntonet, C.G.; and Turias, I. (2010). Improved likelihood ratio test based voice activity detector applied to speech recognition. *Speech Communication*, 52(7-8), 664-677.
10. Chin, S.W.; Seng, K.P.; Li-Minn, A.; and Lim, K.H. (2010). Improved voice activity detection for speech recognition system. *International Computer Symposium (ICS)* Tainan, Taiwan, 518-523.
11. Hoffman, M.W.; Li, Z.; and Khataniar, D. (2001). GSC-based spatial voice activity detection for enhanced speech coding in the presence of competing speech, *IEEE Transactions on Speech and Audio Processing*, 9(2),175-178.
12. Erinnoviar, S.M.; and Hayashi, S. (2000). Speech enhancement with voice activity detection in sub bands. *Bulletin of Science and Engineering*, Takushoku University, 7, 49-54.
13. Srinivasan, K.; and Gersho, A. (1993) Voice activity detection for cellular networks. *Proceedings of IEEE Workshop on Speech Coding for Telecommunications*, 85-86.