Wavelet Region implanting watermark upgrades the security framework in Digital Speech Watermarking

*Dr.S.China Venkateswarlu 1, Dr.Naluguru Udaya Kumar2, Annam Karthik3

1Professor in Dept.of ECE, Institute of Aeronautical Engineering (Autonomous), JNTUH Hyderabad, India
2Associate Professor in Dept. of ECE, Marri Laxma Reddy Institute of Technology and Management (Autonomous), JNTUH Hyderabad, India
3Assistant Professor in Dept. of ECE, Institute of Aeronautical Engineering (Autonomous) JNTUH Hyderabad, India

1cvenkateswarlus@gmail.com, 2joyudaya@gmail.com, 3karthik011190@gmail.com

Abstract: The progressions in innovation helped us to use biometric characteristics to verify the people without having them physically. To limit unapproved people and to encourage approved people different human characteristics are used. Face, unique marks, retina, iris, and DNA are some of the biometrics which is used most often to identify people. For additional security in-person authentication and identification systems, a combination of this biometrics can be utilized. Voice can also be used as a biometric similar to other biometrics. Special equipment and computer systems are required to separate this biometrics. This work helps to enhance security by upgrading the modules and embedding the watermark in the speech signal, in-person verification systems. Speakers are validated based on the watermark present in the speech signal. Energy calculations are performed for detail coefficients to select the coefficients where the watermark has to be implanted. The coefficients with less energy are selected for watermark embedding. Inverse discrete transform is applied on approximation and detail coefficients to produce the watermarked speech. The exhibition of the work is assessed by utilizing the subjective and objective measurements such as peak signal to noise ratio, bit error rate, and perceptual evaluation of speech quality. Peak signal to noise ratio is calculated between unique watermark and separated watermark and, unique and watermarked speech. An arrangement of speech articulations of the speaker during the training stage is used to develop the models for the person authentication system. During testing, structures are removed after original and watermarked speech articulations and are used towards calculate the correctness of the system.

Keywords: Authentication; Approximation coefficients; Biometrics; Bit error rate; Detail coefficients; Discrete wavelet transform; Peak signal to noise ratio; Watermarked speech; Perceptual evaluation speech quality

1 Introduction
Digital speech watermarking is one of the methods used on behalf of security purposes. In recent times, security has been an important aspect in many fields. Biometric authentication is one such security system which is welcomed in many fields. With the advancement in technology biometrics are used to authenticate persons without having them physically. Unique biological characteristics are used as biometrics to verify a person. Biometric authentication helps to limit the unauthorized persons from performing an intended service. Most commonly used human traits to authenticate persons are face, fingerprint, iris, retina, and DNA. Voice can also be used as a biometric.

In voice biometrics, digital watermark is used to enhance the security. [18]Watermark is embedded in the speech signal to obtain watermarked speech. A text or image can be used as a watermark. Watermark is then extracted from the watermarked speech to obtain the original speech. Objective and subjective measures such as PESQ (Perceptual Evaluation of Speech Quality) and PSNR(Peak Signal to Noise Ratio) respectively are calculated to analyse the efficiency of Innovative also removed watermark efficiency.
2. Methodology

2.1 Discrete wavelet transform and inverse discrete wavelet transform:
A discrete wavelet transformation (DWT) requires the usage of a wavelet transformation that utilizes a discrete system of wavelet scales also meanings that conform to those unique laws. This transform the signal interested in a generally symmetrical usual of wavelets, which is the theory distinction between the continuous wavelet transformations (CWT) or its application on behalf of the discrete time sequence here and there named the discrete time constant [4] wavelet transformation (DT-CWT). The wavelet can be created beginning a mounting capacity that depicts its scaling possessions.

DWT is ideal for de noising and compressing of signals. This helps to represent the signal with fewer coefficients resulting in sparser representation. It gives same number of coefficients as the length of the input signal provided. So, it requires less memory. [15] DWT is similar to comparing discrete wavelets with multi-rate filter bands. Analysis of signals at narrower sub bands at different resolutions is possible using DWT.

\[
W_{\phi}(j_0, k) = \frac{1}{\sqrt{M}} \sum_n s(n) \phi_{j_0,k}(n) \quad (1)
\]

\[
W_{\phi}(j, k) = \frac{1}{\sqrt{M}} \sum_n s(n) \phi_{j,k}(n) \quad (2)
\]

where \( j \geq j_0 \) and \( s(n), \phi_{j_0,k}(n), \phi_{j,k}(n) \) are functions of discrete variables.

2.2 Proposed speech watermarking:

Figure 1: Architecture of digital speech watermarking

Figure 2: Watermark generation
2.3 Authentication of a person:
Person authentication framework includes the utilization of training stage where a collection of training speech articulations of the speaker is changed over into a group of features and hence the testing stage and the templates where a group of speech test articulations are occupied after the watermarked speech articulations and to validate the speakers these test articulations are applied to the formants. In areas where security is without a doubt is required, these [19] frameworks are used. Speech utterances are looked over the "NOIZEUS" database where the speaker's utterances are combined with noise at different SNR levels. Mel frequency perceptual linear predictive cepstrum (MFPLPC) features that portray the fundamental attributes of the talks of the speaker are utilized here. Preparation calculation is done by applying the extricated features which in turn produces the templates for each individual speaker. In testing procedure, closeness of templates and test utterances are tested. Components collected from the test joints are included in the 19 predefined models of both speakers and the contrast is produced on the basis of the degree of proximity. You may classify the speakers in this way.

2.4 Feature extraction:
The extraction of features depends upon 2 processes. They are: 1. Vocal tract modelling and 2. Homomorphic filtering of speech. In the first process, when an excitation is passed through a filter speech is produced. The response from the filter effects the vocal track present during the excitation. In the second process, convolution is done for vocal tract response and excitation by adding the logarithms of their transforms. By applying linear filtering vocal tract is separated from the excitation. The differences in the speeches uttered by different speakers are indicated using bandwidth and formant locations. The features utilized for comparing should possess low intra-class variations and high inter-class variations, they should represent high efficiency in identifying persons where speech can be recycled by way of a biometric on behalf of speaker verification and, they must be good enough to recognize any pattern. MFPLPC discourse investigation technique gives insights regarding the perceptual highlights extraction with channels divided into model scale. The suggested approach involves the preparation process and the classification step for authentication speakers by adding the features of the speakers' innovative / watermarked speech to the models, and the organization is achieved by measuring the difference between the test features and the models. At this moment, a watermark is transmitted through the ideal channel now a speech signal popular the wavelet region also a watermarked speech signal remains generated in the transmission adjacent and the watermarked signal. Within the processor, watermarked speech is permitted to be decomposed by wavelets into [5] calculation also information constants, also a watermark remains derived as of the description constants . Watermark is extricated from the watermarked speech and PSNR is determined between the original and extricated watermark. Additional analysis on the singular confirmation is performed by eliminating the characteristics of the watermarked expression and adding them to the settings, i.e. one of the form speaker models while depending on the presentation metric for the specific display procedure; the person is verified through the course of action of the enrolled speakers.

3. Speaker authentication technique:
Vector quantization code blocks are made for each and every class by extracting the features from training and test information. The extracted characters are tested against all tests of different templates to determine the class of that particular articulation from the classes chosen. Training information for every speaker are modelled by combining eight discourse articulations of that particular speaker. To create templates or prototypes, characteristics are extracted from combined discourse articulations that are intended for 22
training and they are applied to training algorithms. Initial discourse signal is transferred through a pre-emphasis block followed by a frame block to extract the features. Discourse signal is translated into 16 msec overlapping frames for each segment with an overlapping of 8 msec. Hamming window which is ideal for discourse frameworks is used to window every frame so that MFPLPC features can be extracted. By utilizing K-means bundling procedure, the characteristics are applied to vector quantization code blocks to generate code book for the speakers. M clusters are mapped to L training vectors in this procedure. For creating reference models for the speakers, every characteristic vector is properly normalized before passing it as an input. Test data considered to be a first / decoded speech utterance referring to speakers for research. When the characteristics from the test expressions are applied to the reference layouts relating to the speaker, evaluation of the speaker identification verification framework is accomplished. The minimum distance between each test vector and the center of each bundle is found out in the testing mechanism. For each speaker model, the average of minimum distances is calculated. The speaker model which has minimum of averages is best fitted for test articulation.

4. Results

Many biometric features are used for security enhancement in communication systems. Face, iris, fingerprints, voice and hand are used as biometric feature so far. Every biometric feature has its advantages and disadvantages. Comparison of these biometric features can be made based on cost effectiveness, complexity and match rates. Out of all these biometric recognition systems speech is one of the significant way of communication which reveals a lot of information about the speaker’s identity. Speaker recognition system involves speaker authentication and speaker identification. The main problem in using speaker recognition systems is that they might undergo spoofing attacks. Spoofing attacks reduce the performance of the system and the security cannot be maintained. Various spoofing attacks such as speech synthesis, replay attacks, voice conversion and, impersonation can be done on speaker recognition systems. The evaluation of performance of the speaker authentication is made by calculating subjective measures and objective measures. Subjective measures such as PESQ and PSNR are calculated. Objective measure such as SNR is calculated.

![Figure 4: Approximation and detail coefficients](image)

Many biometric features are used for security enhancement in communication systems. Face, iris, fingerprints, voice and hand are used as biometric feature so far. Every biometric feature has its advantages and disadvantages. Comparison of these biometric features can be made based on cost effectiveness, complexity and match rates.

Out of all these biometric recognition systems speech is one of the significant way of communication which reveals a lot of information about the speaker’s identity. [11] Speaker recognition system involves speaker authentication and speaker identification. The main problem in using speaker recognition systems is that they might undergo spoofing attacks. Spoofing attacks reduce the performance of the system and the security cannot be maintained. Various spoofing attacks such as speech synthesis, replay attacks, voice conversion and, impersonation can be done on speaker recognition systems.
Table 4.1. Effect of various spoofing attacks based on different studies

| Type of attack     | Countermeasure | Practicality | Vulnerability |
|--------------------|----------------|--------------|---------------|
| Replay attacks     | Low            | High         | High          |
| Speech synthesis   | Medium         | Medium to high| High          |
| Voice conversion   | Medium         | Medium to high| High          |
| Impersonation      | Non-existent   | Low          | Low           |

The evaluation of performance of the speaker authentication is made by calculating subjective measures and objective measures. Subjective measures such as PESQ and PSNR are calculated. Objective measure such as SNR is calculated.

![Figure 4.1: Approximation and detail coefficients of speaker-1](image)

SNR is high for detail coefficients compared to approximation coefficients. Euclidean distance is more at level 1 of approximation coefficients.

Calculation of coefficients, SNR and Euclidean norm at different levels for speaker-2

**Performance:**

|                  |               |
|------------------|---------------|
| Epoch            | 1000 iterations |
| Performance(MSE)| 0.00          |
| Gradient         | 1.00e-07      |
| Mu               | 1.00e+10      |
| Effective # Param| 0.00          |
| Sum squared Param| 0             |
Figure 4.2.: Approximation and detail coefficients of speaker-3

Figure 4.3: Best training performance plot
The performance of the system is evaluated by calculating mean square error (MSE) and regression. MSE is zero for the clean speech.

Regression:

- Training: $R=1$ \quad Output$=1^{\ast}$Target+1.3e-06
- Test: $R=0.47701$ \quad Output$=-28^{\ast}$Target+-2.8
- All: $R=1$ \quad Output$=1^{\ast}$Target+-0.57

5. Conclusion and Future scope

This work talks about the individual validation framework utilizing speech as a biometric looking at the framework utilizing unique and watermarked speech articulations. This suggested validation system of speakers utilizing exceptional verbalizations of speech is performed by applying the features excluded from the collection of articulations to the ready to render designs on behalf of each speaker also by relating structures removed after examination descriptions towards altogether formats also by selecting the finest version composed of the examination articulation dependent happening the smallest division classifier.
Discourse waterline remains similarly changed over addicted to M diagrams through M tests trendy every edge also therefore changed addicted to twofold characteristics. Thusly, a twofold speech/picture watermark is made. A remarkable game plan of articulations considered for planning is connected, to begin with, diagrams, and disintegrated into estimation also feature constants. Implanting of the watermarked speech remains done popular the detail coefficients of the picked squares. To get the watermarked speech, estimate and switched detail coefficients are tested and filtered up. Attributes remain removed after the watermarked speech indication also further towards all speakers’ designs also are subject to section figuring; a speaker is proclaimed. Watermark remains expelled beginning the element constants resulting from the decay of the watermarked speech in the wavelet region. Implanting watermark upgrades the security of the framework. The efficiency of the framework gets improved. Robustness of the speaker recognition system against intentional and unintentional spoofing attacks is enhanced. The bit error rate and mean square error are also reduced. Voice biometrics can be utilized in different fields. They can be installed in places, where the fingerprint biometric systems are not in use. Speaker recognition technology provides a platform to communicate credibly. Digital speech watermarking enhances the security in these speaker recognition systems.

References
[1] Das, R. K., Jelil, S., & Prasanna, S. M. (2017). Development of multilevel speech based person authentication system. Journal of Signal Processing Systems, 88(3), 259–271.
[2] A. Revathi, N. Sasikaladevi, C. Jeyalakshmi (2018). Digital speech watermarking to enhance the security using speech as a biometric for person authentication. Springer Science+Business Media, LLC, part of Springer Nature 2018.
[3] Desai, N., & Tahilramani, N. (2016). Digital speech watermarking for authenticity of speaker in speaker recognition system. In 2016 international conference on micro-electronics and telecommunication engineering (ICMETE) (pp. 105–109). IEEE.
[4] Nematollahi, M. A., Al-Haddad, S. A., Doraisamy, S., & Ranjbari, M. (2014). Digital speech watermarking for anti-spoofing attack in speaker recognition. In 2014 IEEE on region 10 symposium (pp. 476–479). IEEE.
[5] Nematollahi, M. A., Gamboa-Rosales, H., Akhaee, M. A., & Al-Haddad, S. A. (2015). Robust digital speech watermarking for online speaker recognition. Mathematical Problems in Engineering.
[6] Dey, S., Barman, S., Bhukya, R. K., Das, R. K., Haris, B. C., Prasanna, S. R., & Sinha, R. (2014). Speech biometric based attendance system. In 2014 twentieth national conference on communications (NCC) (pp. 1–6). IEEE.
[7] Desai, N. G., & Tahilramani, N. V. (2016). Speaker recognition system using watermark technology for anti-spoofing attack: A Review. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, 4(4), 152–156.
[8] P Sivaraman, and P Prem, “PR Controller Design and Stability Analysis of Single Stage T-Source Inverter Based Solar PV System Journal of Chinese Institute of Engineers Vol.40, No. 3 pp. 235 - 245 APR 2017 [9] Hermansky, H., & Morgan, N. (1994). RASTA processing of speech. IEEE Transactions on Speech and Audio Processing, 2(4), 578–589.
[10] Nematollahi, M. A., Gamboa-Rosales, H., Martinez-Ruiz, F. J., Jose, I., AlHaddad, S. A., & Esmaeilpour, M. (2017). Multi-factor authentication model based on multipurpose speech watermarking and online speaker recognition. Multimedia Tools and Applications, 76(5), 7251–7281.
[11] Rani, R., & Sachdeva, R. (2016). Genetic algorithm using speech and signature of biometrics.
[12] Revathi, A., & Venkataramani, Y. (2011) Speaker independent continuous speech and isolated digit recognition using VQ and HMM. In 2011 International conference on communications and signal processing (ICCCSP) (pp. 198–202). IEEE.
[13] S.China Venkateswarlu, N. Udaya Kumar, K. Chaitanya, A Usha Sree “Effect of feature normalization objective improvement of over Noisy Single-channel Speech Enhancement with Neural Networks”, TEST Engineering & Management, Volume- 83, Issue: May-June 2020, ISSN: 0193-4120, pp.8121-8131, Published By: Published by: The Mattingley Publishing Co., Inc.,
[14] S.China Venkateswarlu, A.Karthik and K.Naveen Kumar “Implementation of Area optimized Low power Multiplication and Accumulation”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, pp. 2928-2932, Volume-9, Issue-1, November 2019. Published By: Blue Eyes Intelligence Engineering & Sciences Publication, Retrieval Number: A9110119119/2019©BEIESP DOI: 10.35940/ijitee.A9110.119119.

[15] S.China Venkateswarlu, A.Karthik and K.Naveen Kumar “Performance on Speech Enhancement in Objective Quality Measures Using Hybrid Wavelet Thresholding”, International Journal of Engineering and Advanced Technology (IJEAT), Volume 6, Issue 8(2019), pp.3523-3533 published on August 2019. DOI: 10.35940/ijeat.F9343.088619

[16] S.China Venkateswarlu, A.Karthik, K. Naveen Kumar, (2019).” Performance on Speech Enhancement Objective Quality Measures Using Hybrid Wavelet Thresholding”, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8, Issue-6, August 2019.

[17] S.China Venkateswarlu, A.Subba Rami Reddy and K.Satya Prasad “Speech Enhancement in Terms of Objective Quality Measures with Effect of Adjustable Window Shape Parameters”, WJTER, Volume 2, Issue 1(2017) 347-359 published on December 2017.

[18] S.China Venkateswarlu, A.Subba Rami Reddy and K.Satya Prasad “Speech Enhancement in Terms of Objective Quality Measures with Effect of Adjustable Window Shape Parameter on Speech Enhancement Techniques”, Cromejournals based – Journal of Signal and Image Research published on July 2016.

[19] S.China Venkateswarlu, A.Subba Rami Reddy & K.Satya Prasad, “Speech Enhancement in terms of Objective Quality Measures Based on Wavelet Hybrid Thresholding the Multi-taper Spectrum”, Published in IJAREEIE, Volume 5, Issue 1, pp.201-219, January 2016, DOI:10.15662/IJAREEIE.2015.0501036.