European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 07, Issue 08, 2020

Speech Forgery Detectionof Framed Sentences In Audio Recordings Using DTW

[1]KasiprasadMannepalli, [2]V.SubbaRamaiah, [3]K.Raghu,

[1]Associate Professor, Department of ECE, koneruLakshmaiah Education Foundation, Vaddeswaram, Guntur(Dist), Vijayawada, A.P, India. Email: mkasiprasad@gmail.com
[2]Assistant Professor, Department of CSE,MAHATMA GANDHI INSTITUTE OF TECHNOLOGY,Hyderabad, Telangana, India.Email: vsubbaramaiah_cse@mgit.ac.in
[3]Assistant Professor, Department of ECE,MAHATMA GANDHI INSTITUTE OF TECHNOLOGY,Hyderabad, Telangana, India.Email: raghukasula@mgit.ac.in

Abstract:
Our main objective is to detect the speech forgery that is formed by framing sentences. These sentences can act as a crucial evidence in courts and in audio forensics and in NLP. In this these sentences are formed with the help of Telugu alphabets which are audio recorded. The alphabets act as a source and basics to form these sentences. The procedure of forming these sentences is as follows from alphabets to words and words to sentences by concatenation. After forming these sentences with help of MFCC which is feature extraction technique used in audio and speech data extraction. After the extraction we are using dynamic time warping technique to estimate the forgery in the audio which is formed from framing sentences. This forgery detection of framed sentences can be further be used in many ways such as in NLP and speech to text conversion applications in audio forensics and some of the new generation applications. The proposed method is able to differentiate the forgery by comparison from original and forged frame sentences. This is done and implemented with help of MatLab tool.

Key words: MFCC, DTW, Audio forensics, Speech forgeries, Matlab, ZCR

1. INTRODUCTION

Audio and Speech recordings are used widely as a digital evidence in the courts and audio forensics [1] very abundantly now-a-days. Technology is increasing very much that forging of the audio is very simple now-a-days with help of simple software tools which are available. Even though with help of technology many methods and techniques have been developed to detect this forgery in the audio but due to some powerful post processing algorithms the detection is not very efficient compared to what we have at present. Even though good number of detections have been done with what we have actually, this research is our continuation of our previous work which is copy and move detection in audio recordings[1]. There are many other techniques which are present like audio splicing, image copy and move detection [2], audio tampering detection [3], audio compression devices etc. These things have gained a lot of attention towards forensics that are working towards forgery detection. Many discoveries have made mentioned as above that has made many researchers enthusiastic towards them. However only few techniques and methods have been published in the scientific research work. As we already mentioned this is our continuation to our previous work. we have come up with another interesting detection technique that will be very useful in forgery detection methods. One of its challenging issue is that to find its strong features so our method may be robust against several methods and techniques Also, to obtain the maximum output with minimum error so that it is very helpful in detecting the forgery very accurately. The biomedical signals analysis can be better performed when good features extracted from the signals[2,3] As speech is not as stationary signal, it needs extract the features by windowing the signal. In much speech processing researches temporal features [4]. Auditory features, like MFCC features [5], pitch chroma[6], spectral flux[7] and tonal power ratio[8] were the feature sets. To classify the pattern of signals for a particular application, classification techniques need to be used. There are many different such techniques

are reported for different applications [9-12]. Based on the data acquisition method a signal de-noising may also be incorporated [13-16]. Transform techniques are useful to obtain the enhanced image in the application like MRI and CT and various assessable analyses are reported on the MRI and CT image [17-19]. Basing on the pitch and LPC based formants, the deviations in the speech segments can be traced [20] The rules package of dynamic time warping (DTW) is implemented to measure the differences between each pitch series and process of formation.

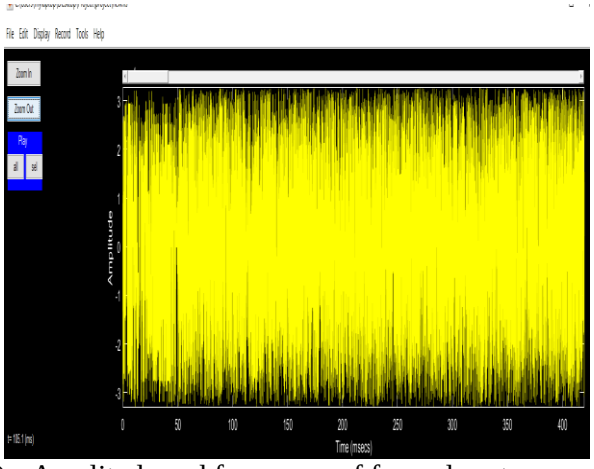
2. LITERATURE SURVEY

Speech Forgeryin Telugu Audio Recordings.

In this part we are going to discuss about the speech forgery detection and some of the methods in detection and analyse the features obtained from the process. The methodology is as follows.

a. features of Speech forgery in Audio

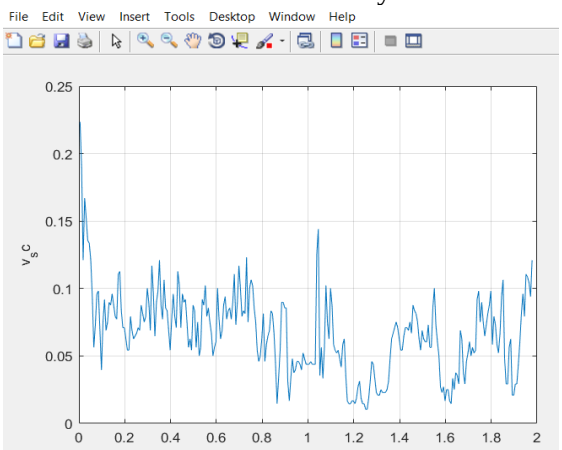
Speech forgery identification is a normally utilized strategy to produce advanced sound chronicles. Any individual can without much of a stretch control the sound chronicle just by replicating the fragment of a sound at one position and sticking a similar portion in other position. The portion of one sound can be glued in the different other sound documents through assistance of numerous product's accessible on the planet. Moreover, with assistance of ground-breaking post-preparing procedures like including commotion in the fragments, inspecting, quantization, separating, packing and so on it is hard to track down or follow the sound sections.



a) Amplitude and frequency of framed sentence

A. Zero Crossing Rate

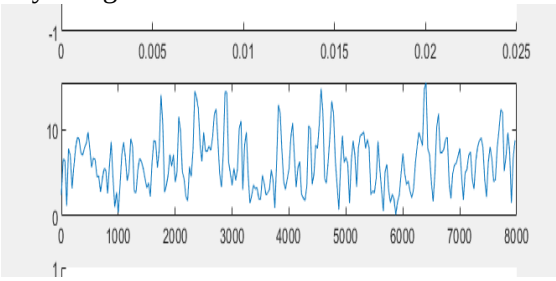
We for the most part utilize zero intersection recurrence in sound acknowledgment and recuperation of music data, which is primary component for arranging percussive sounds. It is the recurrence at which the sign changes from positive to neutral and again from neutral to negative. We can utilize this as a pitch recognition technique and voice identification technique for example, regardless of whether the sound is created by human or machine.



b) Zero crossing rate of the framed sentence

B. MFCC

In sound hearing, we generally use this MFC (cephalic recurrence coefficients) as a transient force scope of a sound. MFCCs are just cephalic coefficients of mel recurrence that are coming about as a result of the cephalic depiction of a sound fasten. The qualification between the cepstrum and the cepstrum of the mel repeat is that in the MFC, the repeat bunches are equidistant on the mel scale, which is closer to the response of the human sound-related system diverged from the legitimately scattered repeat bunches used in the conventional hedgehog. This contortion in repeat it can allow an depiction of the sound, for example in sound weight we use this methodology. These coefficients turn the whole sound sign into shorter edges. We are going to calculate the power spectrum for each and every cell frame that is present in the audio and remove the unvoiced data by using mel filter bank.



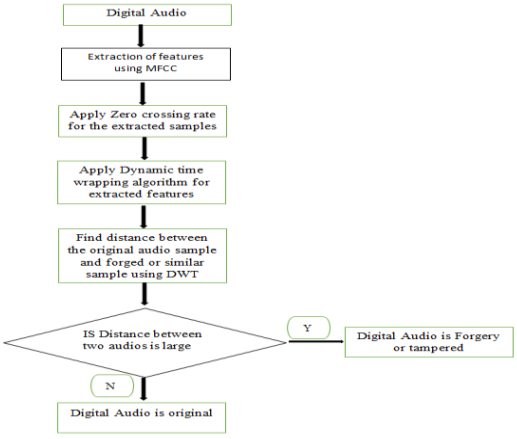
c) MFCC extraction of the framed sentence

3. COLLECTION OF DATA

The collection of data is an important process which decides the whole project. Our team has gathered the data samples of the Telugu alphabets recordings which are of 56 characters using cell recorder through-out the process maintain a bit rate of 150kb/sec in a noiseless and closed environment. Our team also collected the original voice of the speaker to identify the authenticity between original and framed sentences. We have also taken precaution of pre-processing the obtained data so that any noise present in the audio may be detected and removed. We have also applied feature extraction methods to the following data.

4. METHODOLOGY

In this Research project we are using agile method to detect the speech forgery that is done in speech segments. In this we are primarily using feature extraction techniques to extract the features from the audio recordings. Some of the feature extraction methods used in our project are Zero crossing rate and MFC (melcepstrum coefficients). After the extraction of the features we will be using DTW algorithm to mathematically calculate the distances and cepstrum frequencies of original and framed sentences. Then after computation of all the features we will classify the forged framed sentences.



D) Flow chart for the given methodology.

4.1. Extracting the feature's in the audio recording's

In this segment we will be discussing about the extraction methodology. We will be extracting data samples which are nothing but features from the audio recording by using zero crossing rate and MFCC. We will be using both the features and compute the feature similarity between them so, that we could obtain the unique features present in both. The more the features the better the result will be obtained. The zero crossing rate values helps us in such a way that which are present between neutral and negative can be neglected as there will be no audio and it will be stationary.

We will use the MFCC which are mel-frequency coefficients. The coeffcients obtained from the process helps us in calculating the variance and standard deviation from the power spectrum of the audio. This method is done to multiple samples and each and every features are unique in their respective way.

4.2. Similarities of feature extraction and dynamic time warping

The features obtained from the MFCC and ZCR are further given to DTW. The features are separated and each feature is computed individually and the obtained results are stored separately. DTW is a kind of algorithm which measures or computes the distances between two data sequences. The distance as a final result will help us to classify between framed sentences and original sentences.

5. RESULTS AND DISCUSSIONS

The DTW is a kind of algorithm which helps to measures the equidistance between two or more Subsequent data samples and classify into them. In this we have considered the distance between original and forged speech that is formed from framed segments and between framed-framed audio and original-original audio as shown in table.

DTW DISTANCE VALUES AND MAXIMUM DIFFERENCE IN FRAMED SENTENCES

DTW distance values	MFC1	MFC2	MFC3	MFC4	MFC5	MFC6	MFC7	MFC8	MFC9	MFC10
Distance between two Original segments	0.037488	0.016991	0.00629	0.09972	0.05779	0.01999	0.07447	0.03088	0.02744	0.07555
Distance between original and framed segments	0.227033	0.240288	0.35188	0.14088	0.29133	0.17055	0.15300	22.6240	0.47400	0.30166
Distance between two framed segments	0.1666	0.0700	0.018	0.047	0.104	0.131	0.114	1.148	0.140	0.070

	26	5	7	92	8	9	5	8	8	5

A. Computing using dynamic time Warping

As we can see that the obtained features from the MFCC and ZCR are taken into consideration and DTW is applied to those features. From the table it is vivid that the distance between two original audio segments is more than the distance between two framed audio recordings. If we compare the results it is as follows the distance between original to framed audio recordings is much more than any other as per shown in the table. The maximum difference is taken into consideration as we can see clearly. The DTW algorithm really helped us in working with data and is efficient in some way.

6. CONCLUSION

In the current project, we have already projected and investigated the speech forgery detection of framed sentences. The experimental and theoreticalresults had made us to move towards our method with help of the features which are extracted from the data using MFC and ZCR. The DTW method helped in calculating the remoteness and nearness between framed and original. Our proposed method can be effective and smoothly used in detecting and classifying whether the speech segment which has been done is framed for forgery or not based on the experimental analysis. However, the proposed method may have some limits, in the future work, in order to advance the accuracy of the detection even when the forgery is made with some of the powerful detection hiding techniques which are used in the audio recordings that can prevent the change in pitch and formant levels and regularize the data. The method which we have proposed may not get accuracy under such conditions but, considering it as a initial step towards our project we can improve our methodology and indulge new techniques to the above process so that its accuracy can be improved under such conditions in our future work.

7. REFERENCE

1. Q. Yan, R. Yang and J. Huang, "Copy-move detection of audio recording with pitch similarity", *IEEE Int. Conf. Acoust. Speech Signal Process.*, no. 61202497, pp. 1782-1786, 2015.
2. HariPriya, D., Sastry, A. S. C. S., &Rao, K. S. (2016). Low power cmos circuit design for R wave detection and shaping in ECG. ARPN Journal of Engineering and Applied Sciences, 11(24), 14491–14496.
3. Gattim, N. K., Pallerla, S. R., Bojja, P., Reddy, T. P. K., Chowdary, V. N., Dhiraj, V., &Ahammad, S. H. (2019). Plant leaf disease detection using SVM technique. *International Journal of Emerging Trends in Engineering Research*, *7*(11), 634–637. https://doi.org/10.30534/ijeter/2019/367112019
4. .Mannepalli, K., Sastry, P.N., Suman, M.,"Accent recognition system using deep belief networks for telugu speech signals", Advances in Intelligent Systems and Computing, 515, 2017, pp. 99-105. https://doi.org/10.1007/978-981-10-3153-3_10
5. Mannepalli, K., Sastry, P. N., & Rajesh, V. (2015). Accent detection of Telugu speech using prosodic and formant features. In *International Conference on Signal Processing and Communication Engineering Systems - Proceedings of SPACES 2015, in Association with IEEE* (pp. 318–322). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/SPACES.2015.7058274
6. Mannepalli, K., Sastry, P. N., &Suman, M. (2016). MFCC-GMM based accent recognition system for Telugu speech signals. *International Journal of Speech Technology*, *19*(1),87–93. https://doi.org/10.1007/s10772-015-9328-y
7. Mannepalli, K., Sastry, P. N., &Suman, M. (2018). Emotion recognition in speech signals using optimization based multi-SVNN classifier. Journal of King Saud University-Computer and Information Sciences.

a. https://doi.org/10.1016/j.jksuci.2018.11.012

8. Mannepalli, K., Sastry, P. N., &Suman, M. (2016). FDBN: Design and development of Fractional Deep Belief Networks for speaker emotion recognition. , International Journal of Speech Technology, 19 (4), 779-790. https://doi.org/10.1007/s10772-016-9368-y
9. Srinivasa Reddy, S., &Suman, M. (2018). Microaneurysm extraction with contrast enhancement using deep neural network. *Journal of Advanced Research in Dynamical and Control Systems*, *10*(11), 313–320.
10. Reddy, S. S., Suman, M., &Prakash, K. N. (2018). Micro aneurysms detection using artificial neural networks. *International Journal of Engineering and Technology(UAE)*, *7*(4),3026–3029. https://doi.org/10.14419/ijet.v7i4.14895
11. Bojja P., Sanam N., Design and development of artificial intelligence system for weather forecasting using soft computing techniques, ARPN Journal of Engineering and Applied Sciences, Vol:12, issue:3, 2017, pp: 685-689, ISSN: 18196608.
12. Vallabhaneni R.B., Rajesh V., Brain tumor detection using mean shift clustering and glcm features with edge adaptive total variation denoising technique, ARPN Journal of Engineering and Applied Sciences, Vol:12, issue:3, 2017,pp: 666-671, ISSN: 18196608.
13. Gattim, N. K., Rajesh, V., Partheepan, R., Karunakaran, S., & Reddy, K. N. (2017). Multimodal image fusion using curvelet and genetic algorithm. *Journal of Scientific and Industrial Research*, *76*(11), 694–696.
14. Bhavana, D., & Rajesh, V. (2016). A new pixel level image fusion method based on genetic algorithm. *Indian Journal of Science and Technology*, *9*(45),1–8. https://doi.org/10.17485/ijst/2016/v9i45/76691
15. Bhavana, D., Rajesh, V., & Kumar, K. K. (2016). Implementation of plateau histogram equalization technique on thermal images. *Indian Journal of Science and Technology*, *9*(32), 1–4. https://doi.org/10.17485/ijst/2016/v9i32/80562
16. Bhavana, D., Rajesh, V., &KoteswaraRao, C. H. (2016). Multispectral image fusion using integrated wavelets and principal component analysis. *International Journal of Control Theory and Applications*, *9*(34), 737–743.
17. Revathi, B., Naveen Kishore, G., &Dheeraj, V. (2019). A survey on OCR for Telugu language. *International Journal of Scientific and Technology Research*, *8*(12), 559–562.
18. BenniloFernandes, J et. al, "Fuzzy utilization in speech recognition and its different application", International Journal of Engineering and Advanced Technology (2019), 8 (5 Special Issue 3), pp. 261-266. https://doi.org/10.35940/ijeat.E1058.0785S319
19. BenniloFernandes, J., Sivakannan, S., Prabakaran, N., &Thirugnanam, G. (2018). Reversible image watermarking technique using LCWT and DGT. *International Journal of Engineering and Technology(UAE)*, *7*(1), 42–47. https://doi.org/10.14419/ijet.v7i1.3.9224
20. Q. Yan, S. Member, R. Yang and J. Huang, "Robust Copy – Move Detection of Speech Recording Using Similarities of Pitch and Formant", *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 9, pp. 2331-2341, 2019.
21. Mannepalli, K., Sastry, P. N., &Suman, M. (2017). A novel Adaptive Fractional Deep Belief Networks for speaker emotion recognition. *Alexandria*
22. *Engineering Journal*, *56*(4), 485–497. https://doi.org/10.1016/j.aej.2016.09.002