Fast Speaker Identification using Recursive Word Sample Attributes

K. V. S. Ramachandra Murthy
Department of Electrical & Electronics Engineering, Aditya University

V. Satyanarayana
Department of ECE, Aditya University

https://doi.org/10.5109/7183441

出版情報：Evergreen. 11 (2), pp.1313-1319, 2024-06. 九州大学グリーンテクノロジー研究教育センター
バージョン：

Fast Speaker Identification using Recursive Word Sample Attributes

K. V. S. Ramachandra Murthy[1], V. Satyanarayana[2*]

[1]Department of Electrical & Electronics Engineering, Aditya University, Surampalem, India

[2]Department of ECE, Aditya University, Surampalem, India

Corresponding Author E-mail: vasece_vella@adityauniversity.in

(Received September 14, 2022; Revised April 8, 2024; Accepted June 14, 2024).

Abstract: Voice identification is the process to identify a human, based on his/her voice. In this paper, we propose a voice identification system using a technology of 5-Means Clustering approach on both Bandwidth and Formants Ratio. At first, the PRAA database is trained by these methods, this method has used five clustered for bandwidth and also formants ratio. These clusters are classified on the behalf of fuzzy members, and thereafter each and every cluster of bandwidth and formants ratio are combined in different type of combinations. In the testing section, voice samples taken from speakers are compared with these combinations. These systems make them provided fast voice identification process. The Experiment result of the proposed work is achieved more than 98% of Identification Rate.

Keywords: Bandwidth, Clustering, Feature Extraction, Formants, Fuzzy, Voice Identification.

1. Introduction

The human voice is the one which is the mostly readable biometrics for human identification because of uniqueness. This biometric trait is generated by humans in order to convey the messages. The voice is a kind of vibrations that can be recorded in terms of signal. In real time observation also reveals that this attribute is very much unique for every human being [1, 2, 3, 4). Furthermore, the characteristics of this biometrics also varies while recursive production of the same word. The motive of presented speaker identification system is to find out the presence of registered speaker on the basis of voice. Although the voice is self-sufficient identification under noise free environment, it becomes more complex in complete noise environment. In the voice identification, if noise is present, it must be removed. The basic application of the voice identification is in banking, product buying products, robotics and solving crimes in forensic investigation [5, 6, 7, 8, 9).

The simplified block diagram for the voice recognition system is depicted in Fig. 1, where from the voice samples the training and test voice databases are generated [10, 11). Each voice sample under goes the feature extraction process. Subsequently, the features undergo for classification by which the speech speaker identification is performed [12, 13).

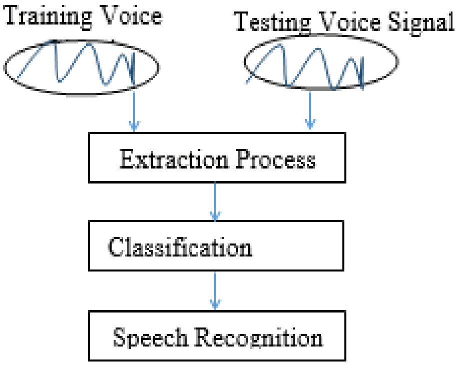

Fig. 1: Voice Recognition System

2. Related Works

Khan et al. voice recognition technology can be classified into two branches; the first branch it is called text-independent and other is called text-dependent methods. In the text-independent environment, where speaker use un supervised learning method. In a text-dependent environment, the speaker's identity is based on his or her speaking voice, that use one or more specific words, passwords, card numbers and etc· [1, 14, 15). Fang introduce a speaker recognition technology; this is based on some phonemes like stops, nasals, and vowels [3, 16). Where vowels have more and more special information for voice identification. In the feature extraction, extracted the voice signals and this is called first step of the voice recognition that showing in Fig. 1 [2, 6, 17)]. It was used to make decisions on how to identify an unknown speaker and to simulate an acoustic spectrograph speech recognition project. In this investigation, they examined five, seven, eighteen, and nineteen concepts in addition to the impact of five different variables on speaker identification. Five parameters are considered: the number of known or unknown speakers in the collection; the test environment type (open or closed); the speech transmission scheme; the context of the speech materials; and the timeliness or non-contemporaneity of the voice input. They discovered two different types of errors: false rejection and incorrect identification (8, 22). When compared to earlier techniques, accuracy rates for both closed collections of isolated words and contemporary voice input have significantly improved. The result for older voice samples is especially relevant for forensic situations since it provides hard evidence of the negative effect on speech recognition. 22 and 23.

There are several challenges in identifying an unknown speaker, the most significant of which are discussed here. In an attempt to discover solutions to these issues, they posed numerous pertinent questions. The question of whether a speaker's formants or pitch change as they age has not yet been addressed [10, 11, 24)]. Researchers aim to find out if the phonemes of a speaker vary with age and health or if they remain constant throughout their lives. However, another crucial question they fail to address is whether the formants of the spectrograms of the voices in disguise differ from those of the real sounds [12, 25)]. This also applies to imitators, who attempt to mimic the target's vocal pitch by matching the spectra of well-known speakers with their own spectrogram. The findings demonstrate that mimics are unable to match the target speaker's frequency position as closely as the original (13, 26). This means that while attempting to identify copycats, factors other than audibility and average pitch frequency should be taken into account. Despite the fact that intonation, speech dynamics, loudness, classic phrases, and dialects are additional factors that significantly influence imitation, spectrograms are not a good way to characterize or explain them [14, 15, 27)].

The main objective of the online voice recognition research experience was to develop computer software that can differentiate between different speech speakers. Our objective was to develop a method that might be used to tackle the "real life" problem of speech recognition using abstract scientific ideas [2)]. The technique was believed to have succeeded in one of its objectives, which was "to be able to distinguish person based on the voice" [15)].

- ➢ A thorough review of the relevant literature on voice identification systems is required.
- ➢ The goal is to find the most effective strategy for voice parameterization by analyzing the several existing approaches.
- ➢ To explore new techniques for improving the voice identification.

3. Fuzzy Clustering Method

There are two type of analysis can be performed on voice signal namely acoustic analysis and spectral analysis. In this paper we used spectral because it is provided a wide range of analysis. Bandwidth and formants are known as the key spectral features of voice signals.

In the section, the PRAA database has been used for training purpose. This database is contained 50 speakers voice of male and also 50 speaker's voice of female. All speakers were asked to provide a set of nine words {Hello, Up, Down, Left, Right, On, Off, turn on, Turn off}.

In this paper, first of all the fuzzy system is applied on bandwidth and formants ratio of all speakers of PRAA database. This fuzzy system used three fuzzy members as μ_A, μ_B and μ_C. After that each speaker are created two clusters, one cluster hold the overall value of bandwidth (B1, B2, B3) and other then stored the value of formants ratio ($f_{r\_1}, f_{r\_2}, f_{r\_3}$) [4].

Fuzzy Member (μ_k) ={ (μ_1,0.05), (μ_2,0.1), (μ_3,0.15)}

$$C_k = \mathrm{O}_{j,k=1}^{j,k=3}(B_k \times \mu_k) \tag{1}$$

$$C_{k+3} = \mathrm{O}_{j,k=1}^{j,k=3}\left(f_{r_j} \times \mu_k\right) \tag{2}$$

The Fig. 2 and Fig. 3 are showing the diagram of fuzzy clustering on bandwidth and formants ratio respectively.

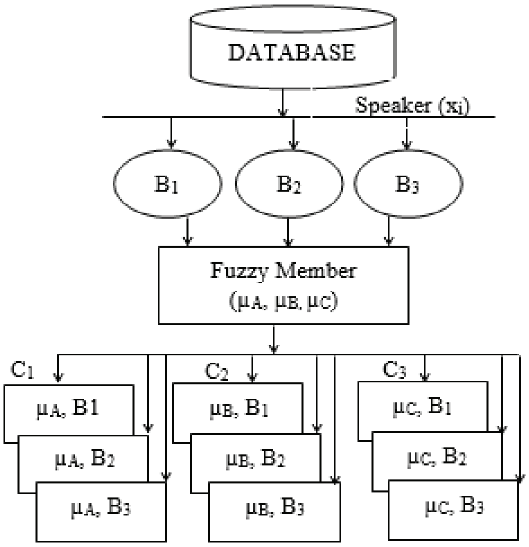

Fig. 2: Fuzzy clustering on bandwidth

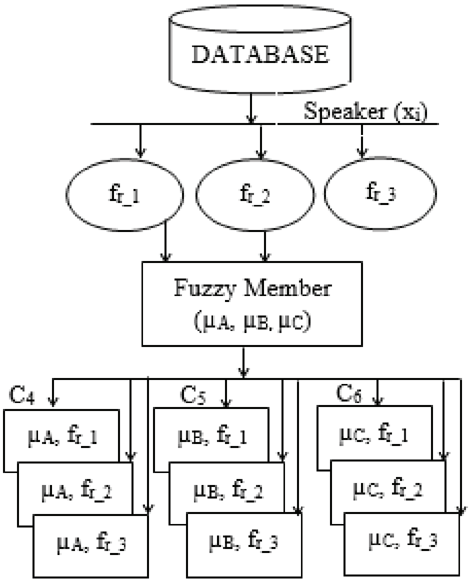

Fig. 3: Fuzzy clustering on formants ratio

After getting the all six clusters of speakers (x_i) is compared with test voice signal(y) for voice identification. If bandwidth clusters (C1, C2, C3) and formants ratio clusters (C4, C5, C6) are equalled tosignal (y) then that means we have identified the speaker otherwise we have not identified the speaker in equation 1.

[{(C1, C2, C3), (C4, C5, C6) of $X_i \approx$ y] → Identified the speaker

Otherwise

[{(C1, C2, C3), (C4, C5, C6) of $X_i \neq$y]→ Not identified the speaker

4. Experimental Results and Discussion

4.1 Dataset description

The proposed scheme was tested on PRAA database in which the samples were taken under constrained environment of limiting the noise. There are 100 samples in the database of 50 male and 50 females in which the occurrence of the recording was done twice.

4.2 Result

In this section, we estimate the performance of the proposed work. So here the value of all six clusters in given below in table 1. Now we have recorded a signal (y) by any device and any environment, here we assume that, noisy environment is presented there $^{5).}$ A signal v (t) of one word is taking as a input in equation 2.

$$y\ (t) = nf\ (t) + n\ (t)\ ,\ 0 \leq t \leq T \qquad (3)$$

Where nf(t) and n(t) are representing noise free signal and noisy signal at any time (t). Now, the PRAAT simulation tools remove the noise of signals, and have calculated the bandwidth and formants that is showed in Fig. 4, Fig. 5 and Table 1.

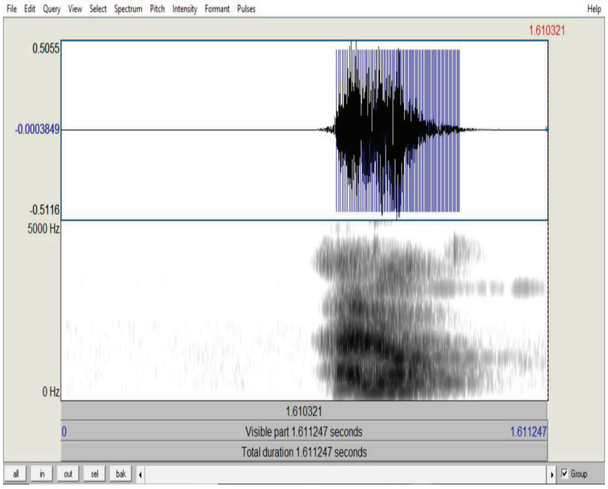
File Edit Query View Select Spectrum Pitch Intensity Formant Pulses
Help
1.610321
0.5055
-0.0003849
-0.5116
5000 Hz
0 Hz
1.610321
0
Visible part 1.611247 seconds
1.611247
Total duration 1.611247 seconds
all in out sel bak
Group

Fig. 4 (a): Spectrogram of “Hello” word

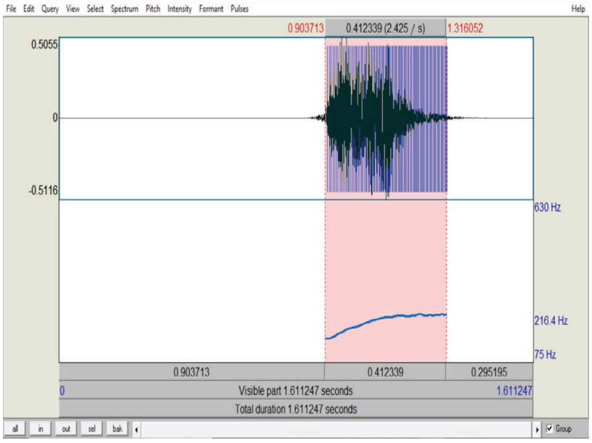

Fig. 4 (b): Bandwidth analysis of “Hello” Word

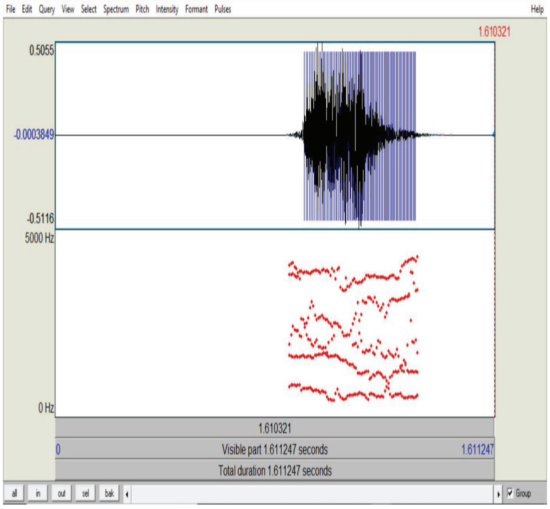
File Edit Query View Select Spectrum Pitch Intensity Formant Pulses
Help
1.610321
0.5055
-0.0003849
-0.5116
5000 Hz
0 Hz
1.610321
0
Visible part 1.611247 seconds
1.611247
Total duration 1.611247 seconds
all in out sel bak
Group

Fig. 4 (c): Formant’s ratio “Hello” word

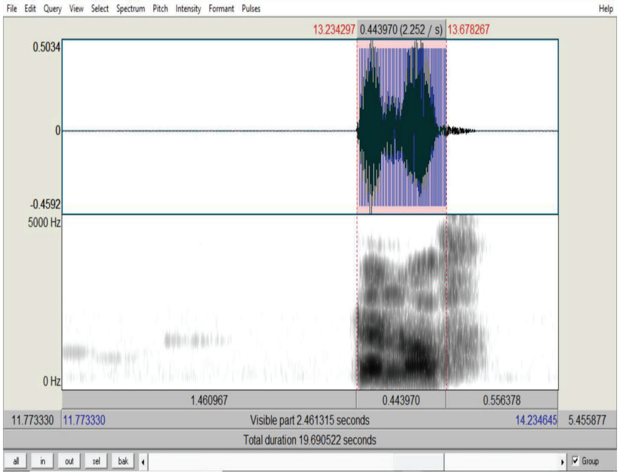

Fig. 5(a): Spectrogram of "Turn on" Word

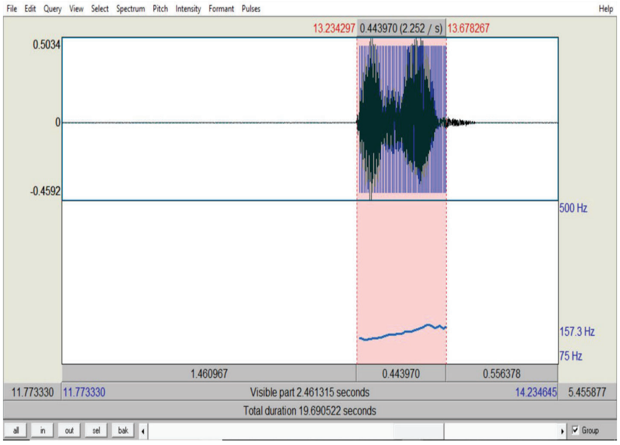

Fig. 5 (b): Bandwidth analysis of "Turn on" Word

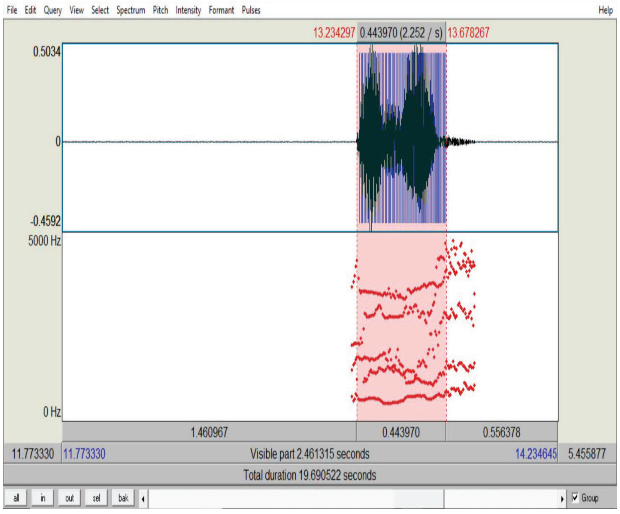

Fig. 5 (c): Formant's ratio "Turn on" word

Figure 4 and 5. Experimental GUI of input signals for testing. For our analysis, we used twenty speakers from each of five languages: Hindi, English, Bengali, Punjabi, and Telugu. It was determined that for each target sentence the twenty most effective reference speakers would be selected based on vectors generated from the 100 reference speakers shown in following tables.

Table 1. Value of input signals for testing

Query	**Bandwidth-B1**	**Bandwidth-B2**	**Bandwidth-B3**
Q_{Hello}	197.972	184.714	2179.974
Q_{Turnon}	41.477	91.245	284.277
	F1/F2	**F1/F3**	**F2/F3**
Q_{Hello}	0.437584	0.302025	0.690212
Q_{Turnon}	0.479079	0.333368	0.69585

Table 2. C1 of Bandwidth

Words	**C1**	Bandwidth value of database speaker
Hello	**188.073**	**207.870**
On	206.778	228.545
Off	73.671	81.426
Left	22.032	24.351
Right	45.372	50.148
Up	152.714	168.789
Down	164.926	182.287
Turn On	**39.403**	**43.550**
Turn Off	111.34	123.06

Table 3. C2 of Bandwidth

Bandwidth value of database speaker (x27)		
Words	**C2**	
Hello	**175.478**	**193.949**
On	582.274	643.566
Off	55.649	61.506
Left	308.560	341.041
Right	198.980	219.925
Up	144.512	159.723
Down	223.679	247.224
Turn On	**86.682**	**95.807**
Turn Off	219.399	242.494

Table 4. C3 of Bandwidth

Words	C3	Bandwidth value of database speaker (x27)
Hello	**2133.396**	**2226.55**
On	470.760	520.313
Off	189.846	209.829
Left	286.472	316.627
Right	611.876	676.284
Up	271.985	300.615
Down	155.310	171.659
Turn On	**270.063**	**298.490**
Turn Off	191.937	212.140

Table 5 C4 of Formants ratio

formants ratio of database speaker (X $_{27}$)		
Words	**C4**	
Hello	**0.41571**	**0.45946**
On	0.32851	0.36309
Off	0.5384	0.59508
Left	0.2858	0.31588
Right	0.26608	0.29408
Up	0.46209	0.51073
Down	0.42753	0.47253
Turn On	**0.40646**	**0.44924**
Turn Off	0.45513	0.50303

Table 6. C5 of Formants ratio

formants ratio of database speaker(x27)		
Words	**C5**	
Hello	**0.28692**	**0.31713**
On	0.18773	0.20749
Off	0.24707	0.27307
Left	0.18845	0.20829
Right	0.20049	0.22159
Up	0.26643	0.29448
Down	0.22099	0.24425
Turn On	**0.2682**	**0.29643**
Turn Off	0.3167	0.35004

Table 7. C6 of Formants ratio

Formants ratio of database speaker (x27)		
Words	**C6**	
Hello	**941.858**	**1151.16**
On	562.514	687.518
Off	1524.51	1863.29
Left	68.9067	84.2193
Right	502.845	614.589
Up	1443.23	1763.95
Down	501.178	612.55
Turn On	**299.568**	**366.138**
Turn Off	817.501	999.167

The above table 2 to table 5 are showed the value of bandwidth and formants ratio respectively of speaker (27) of database, and table 1 is showed the value of bandwidth and formants ratio of input signal (y). .

```
    If {B1, B2, B3 of 'Hello' word} ∈ (C1 OR C2 OR C3)
    If {$f_{r\_1}$, $f_{r\_2}$, $f_{r\_3}$ of 'Hello'} ∈ (C4 OR C5 OR C6)
    Select Cluster Combination ($C_a$, $C_b$)
    End
    End
Otherwise: No Matching
Where $C_a$= (C1, C2, C3) ,$C_b$=(C4, C5, C6)
```

So we can say that, the finally we have achieved cluster combination (C1, C4) on the experimental data. The cluster combination (C1, C4), belongs to 98-100% matching percentage. In Table 8, it is showed the identification rate with respect to cluster combinations.

Table 8. Identification rate with cluster combinations

Clusters Combination	Identification Rate
C1, C4	98-100%
C1, C5	95-98%
C1, C6	80-95%
C2, C4	95-98%
C2, C5	90-95%
C2, C6	80-90%
C3, C4	85-90%
C3, C5	75-85%
C3, C6	65-75%

5. Conclusion

The purpose of proposed study provides an analysis of the formant features for better recognition of the speaker. The observed study of the formant features reveals that these features have low interclass variation for better results and are capable of better discrimination. The proposed scheme employs the formant features of the voice signal and resulting accuracy is of remarkable. This technology is combined of two concepts that make very impressive compared to against the weakness of the conventional voice recognition algorithms. Therefore, proposed work is doing strongly identified voice in text dependent environment.

References

1) Ameen Khan, A., NV Uma Reddy, and Bae, Hyan-Soo, Ho-Jin Lee, and Suk-Gyu Lee. "Voice recognition based on adaptive MFCC and deep learning." *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, (2016). doi: 10.1109/ICIEA. 2016. 7603 830.
2) Fang, Eric, and John N. Gowdy. "New algorithms for improved speaker identification." *International Journal of Biometrics* 5.3-4: 360-369 (2013).
3) Barai, Munim K., et al. "Higher education in private universities in Bangladesh: A model for quality assurance." *Evergreen* 2.2 24-33 (2015) doi:10.5109/1544077.
4) Satya, P. Mohana, et al. "Stripe Noise Removal from Remote Sensing Images." *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)*. IEEE, 2021. doi: 10.1109/ISPCC53510.2021.9609457
5) Yang, Haiya, and Akira Harata. "Design of a Semi-confocal Fluorescence Microscope for Observing Excitation Spectrum of Soluble Molecules Adsorbed at the Air/water Interface." *Evergreen: joint journal of Novel Carbon Resource Sciences & Green Asia Strategy* 2. (2): 1-4 (2015). doi.org/10.5109/1544074
6) Akbar, Lasta Azmillah, et al. "Method development of measuring depth of burn using laser ranging in laboratory scale." *Evergreen* 7.(2) 268-274 (2020):. doi: 10.5109/4055231.
7) Berawi, Mohammed Ali, et al. "Determining the Prioritized Victim of Earthquake Disaster Using Fuzzy Logic and Decision Tree Approach." *Evergreen* 7.2: 246-252 (2020).doi:10.5109/4055227
8) Chauhan, Shailendra Singh, and S. C. Bhaduri. "Structural analysis of a Four-bar linkage mechanism of Prosthetic knee joint using Finite Element Method." *EVERGREEN Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy* 7.02 (2020).doi: 10.5109/4055220.
9) Nandini, A., R. Anil Kumar, and Mahesh K. Singh. "Circuits Based on the Memristor for Fundamental Operations." *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)*. IEEE, 2021. doi: 10.1109/ISPCC53510.2021.9609439.
10) Sharma, Manish, and Rahul Dev. "Review and Preliminary Analysis of Organic Rankine Cycle based on Turbine Inlet Temperature." *Evergreen.* 5: 22-33 (2018) doi: 10.5025/1856985.
11) Anushka, R. L., et al. "Lens less Cameras for Face Detection and Verification." *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)*. IEEE, 2021. doi: doi: 10.1109/ISPCC53510.2021.9609392.
12) Mohd, Nik, et al. "Lattice boltzmann method for free surface impacting on vertical cylinder: A comparison with experimental data." *Evergreen: joint journal of Novel Carbon Resource Sciences & Green Asia Strategy* 4.(2): 28-37 (2017). doi: 10.5109/1929662.
13) Wang, Chaojun, and Fei He. "State clustering of the hot strip rolling process via kernel entropy component analysis and weighted cosine distance." *Entropy* 21.(10): 1019 (2019).doi: 10.3390/e21101019.
14) Balaji, V. Nithin, P. Bala Srinivas, and Mahesh K. Singh. "Neuromorphic advancements architecture design and its implementations technique. "*Materials Today: Proceedings* (2021). doi: 10.1016/ j.matpr. 2021.06.273.
15) Lei, Lei, and She Kun. "Speaker recognition using wavelet cepstral coefficient, i-vector, and cosine distance scoring and its application for forensics." *Journal of Electrical and Computer Engineering* 2016 (2016). doi: 10.1155/ 2016/ 4908412. https://doi.org/10.1007/s11042-023-17368-5
16) Singh, Mahesh K., Narendra Singh, and A. K. Singh. "Speaker's voice characteristics and similarity measurement using Euclidean distances." *2019 International Conference on Signal Processing and Communication (ICSC)*. IEEE, 2019.doi: 10.1109/ICSC 45622. (2019). 8938366.
17) Singh, M. K. (2023). Feature extraction and classification efficiency analysis using machine learning approach for speech signal. Multimedia Tools and Applications, 1-16.
18) Padma, Uppalapati, Samudrala Jagadish, and Mahesh K. Singh. "Recognition of plant's leaf infection by image processing approach."*MaterialsToday: Proceedings* (2021). doi: 10.1016/j.matpr. 2021. 06.297.
19) Singh, Mahesh, Durgesh Nandan, and Sanjeev Kumar. "Statistical Analysis of Lower and Raised Pitch Voice Signal and Its Efficiency Calculation." *Traitement du Signal* 36.(5): 455-461 (2019). doi: 10.18280/ts.360511.
20) Kreyssig, Florian L., and Philip C. Woodland. "Cosine-distance virtual adversarial training for semi-

supervised speaker-discriminative acoustic embeddings." *arXiv preprint arXiv:2008. 03756* (2020). doi: 10.1109/SLT. 2016.7846310.

21) Punyavathi, G., M. Neeladri, and Mahesh K. Singh. "Vehicle tracking and detection techniques using IoT." *Materials Today: Proceedings* (2021). doi: 10.1016/j.matpr. 2021.06.283.
22) Singh, Mahesh K., A. K. Singh, and Narendra Singh. "Multimedia utilization of non-computerized disguised voice and acoustic similarity measurement." *Multimedia Tools and Applications* 79.(47), 35537-35552 (2020). doi: 10.1007/s11042-019-08329-y.
23) Balaji, V. Nithin, P. Bala Srinivas, and Mahesh K. Singh. "Neuromorphic advancements architecture design and its implementations technique." *Materials Today: Proceedings* (2021). doi: 10.1016/j.matpr. 2021.06.273.
24) . Singh, M. K. (2024). Multimedia application for forensic automatic speaker recognition from disguised voices using MFCC feature extraction and classification techniques. Multimedia Tools and Applications, 1-19. https://doi.org/10.1007/s11042-024-18602-4 .
25) Choi, Jonghyun, et al. "Toward sparse coding on cosine distance." *2014 22nd International Conference on Pattern Recognition*. IEEE, (2014). doi: 10.1109/ICPR.2014.757.
26) Singh, Mahesh K., A. K. Singh, and Narendra Singh. "Multimedia analysis for disguised voice and classification efficiency." *Multimedia Tools and Applications* 78.(20): 29395-29411(2019). doi: 10.1007/s11042-018-6718-6.
27) Zou, Bei-ji, and Marie Providence Umugwaneza. "Shape-based trademark retrieval using cosine distance method." *2008 Eighth International Conference on Intelligent Systems Design and Applications*. Vol. 2. IEEE, (2008) doi: 10.1109/ISDA.2008.161.