International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 14 (2015) pp 33976-33980

Cellular Automata Based Noise Removal and Its Application in Speaker Recognition.

Ahmed Sajjad Khan

Department of Electronics and Telecommunication
Anjuman College of Engineering and Technology. Nagpur, India.
email: acetsajjadsir@yahoo.co.in

E.G. Rajan

Professor Of Signal Processing
Professional Member, ACM. Hyderabad, India.
e-mail: rajaneg@yahoo.co.in

Abstract
Noise is unwanted information of speech signal. This kind of noise can occur during speech production systems itself, and can also occur during speech recording. So to enhance the accuracy of speaker recognition process, we have to reduce the noise present in the signal. Most primitive type approaches uses neighbor cell values to replace noisy cell. But it has a big disadvantage that it is applied on the entire cell, corrupted as well as un-corrupted cell. Because of this the speech may lose vital quality such as pitch and formants. In the method proposed we identify the noise and then removes it from the corrupted speech signal based on CA. To illustrate the proposed method, some experiments have been performed on several standard test speeches and compared with popular methods of filtering. The results show that the proposed method relatively has the desirable performance enhancement in speaker recognition. First the concept of CA is introduced in filtering the noise from speech signal, and then accordingly applied to the state of art speaker recognition system to the structure of the neighbor's cells, proposed model and then the experimental results. The improvement of Speaker recognition accuracy after application of CA from 96% to 98% has been obtained in this case.

Keywords-----noise filtering, cellular automata, speech signal, MFCC,neighbourhood structure.

I. INTRODUCTION

Cellular automata is a dynamical system here time and space are discrete. Von neumann [1] and Stephen Wolfran [2] established Cellular automata and it can be used for simulation of evaluation process. In speaker recognition process, two important things which has to be considered is speech production and its recording. During these two processes noise can easily detiorate the original speech signal. The noise can be internal or external. Internal noise can be produced because of illness, fear, phobia etc. External noise is the noise in your surroundings by other people and the environment. External noise sources include atmosphere disturbance example ionospheric effect, lighting, etc,. Cosmic noise such as noise from galaxy, solar noise etc. These noise disturbances produces wrong information in speaker recognition system.

The random disturbances in the speeches are due to noisy channel of transmission, storage at wrong memory locations. This reduces the quality of speech and damages the expression of information for speech effectively. So by filtering the noise the speech signal is made smooth. The different methods of Cellular automata based speech filtering are spatial filtering & frequency domain filtering.

By filtering the noise means, we are basically suppressing the noise signal and here care is taken that the different parameters of speech such as pitch, formants, etc are not affected, that is speech signal is not degraded. Therefore filtering or preprocessing step is applied [3].there are number of cellular automata based algorithms designed. It is found that Median filter removes impulse noises far better than linear filters [4][5]. Other filters are adaptive median filter [6] and progressive switching median [7].

Here we have used filter which are based on the concept of cellular automata, and it removes impulse noises present in the speech signal corrupted by different noises [8] and [9]. The cellular automata algorithm is applicable to speech, and show significant improvements over the performance for speaker recognition [8][9].

II. CELLULAR AUTOMATA

cellular automata concept is based on state of different cells, their neighbors and it is governed by different rules. As time advances at discrete steps, state of new cell is determined by state of present cell along with the present state of its neighboring cell, that too according to a well defined specific rules. The rules of the cellular automata are local and uniform. There are different models of C.A., such as one dimensional C.A, two dimensional C.A and three dimensional C.A. One dimensional C.A with two states consist of line of cells with values "0" and "1".

If we are using One dimensional C.A with n-states, than each cell can have any integer value that is from 0 to n-1. The local rule will determine the status of next cell and the cell values

are updated at each discrete time steps. It means that the local rule will control the evolution of cellular automata.
We can define cellular automata model as follows [1]. A cellular automata of 4 – tuple (L,S,N,F), here L means cells of regular lattice, S means cells of finite state, N means finite number of sets of neighbors which indicates the relative position of a cell in relation to other cells of the given lattice N, and F is a function which will assign a new state to a cell. where $F: S^{|N|} \to S$. There are 256 ($2^8 = 256$) different rules.. To analyze the behavior S. Wolfram developed a convention where he has taken initial condition and developed a method to view the results at a glance after multiple iteration
In one dimensional C.A.'s described, a hierarchy was given to eight possible different patterns with black-black-black (1-1-1) on the far left and white-white white (0-0-0) one the far right. Each combination represents a place in the binary numbering system.
Table. 1 shows a 1Dimensional two state with nearest neighbor CA. The lattice structure is seven cells wide and it is demonstrated at two different time steps that is at t=0 and at t=1. In this example, the local neighborhood configuration of the second cell (1), at time step t = 0 is “0-1-1” (the current values of the first, second and third cells), and from the lookup table we can that this cell will be in state “0.66” at time t = 1.
The local neighborhood configuration of the third cell (1) at given time step t = 0 is “1-1-1” (the current values of the second, third, and fourth cells), and from the given lookup table we can say the cell will be in state “1” at time step t = 1.
Similarly the neighborhood configuration of the sixth cell (1) at given time t = 0 is “1-1-0” (the current values of the fifth, sixth and seventh cells), and from the given lookup table we can say the cell is in state “0.33” at time step t = 1. Similarly the other cells will get updated at time t=1.

Table 1: Initial contribution and after updates

t=0	0	1	1	1	1	1	0
t=1	0.33	0.66	1	1	1	0.33	0.66

III. STRUCTURE OF THE NEIGHBORHOODS

As the speech signal is basically a two dimensional signal, hence we uses 2D cellular automata model. In 2D cellular automata model, we can use triangular lattice structure or square lattice structure or hexagonal lattice structure, but for speech signals the results using square lattice is far convincing as compare to other lattice structures, so here we have used square lattice for analysis. Each speech sample is of 16 bits. Initially we consider original signal at time step t=0.
The two neighborhood structures are, Von Neumann neighborhood structure and Moore neighborhood structure. Apart from this there are number of rules of cellular automata. So for denoising the speech signal we have to find the type of neighborhood structure and depending on this we will select the rules of cellular automata.
The cell present on the right side and on the left side, and the cell present above and below are called Von Neumann neighborhood of the given cell. here we consider the next layer the radius of our defination is 1. As there are four neighboring cells and including itself, the total cells become 5 [2] as given in equ (1)

$$N(I,j) = \{(k,l) \in L: |k-i| + |l-j| \le 1\} \quad (1)$$

where, k is the no. of states for the cell and l is the space of speech signal.
As in Von Neumann neighborhood, there are total 5 cells (1 + 4 neighbours), but in Moore neighborhood there are total 9 cells (1 + 8 neighbours), The cell present on the right side and on the left side, and the cell present above and the cell present below plus 4 diagonal cells. This is given by equation (2)

$$N(I,j) = \{(k,l) \in L: max|k-i|, |l-j| \le 1\} \quad (2)$$

The state of cell at time instant $t+1$ depends on the present state of itself cell plus the state of the cells in the given neighbourhood at time t, this is represented in equation (3)

$$S_{i,j}(t+1) = f\,(S_{i-1,j-1}(t), S_{i+1,j}(t), S_{i-1,j+1}(t), S_{i,j-1}(t),$$

$$S_{i,j}(t), S_{i,j+1}(t), S_{i+1,j-1}(t), S_{i+1,j}(t), S_{i+1,j+1}(t) \quad (3)$$

The main concept of structuring element is used to compare the central amplitude of speech signal with the amplitude of the neighboring cells, as given in the equation (4)

$$SE = Strel('square', w) \quad (4)$$

This creates a square structuring element with a width of w samples.the integer w must be a positive integer scalar as shown in the table 2:

Table 2: Nine neighbourhood structure

i-1, j-1	i-1, j	i-1, j+1
i, j-1	i, j	i, j+1
i+1, j-1	i+1, j	i+1, j+1

IV. METHOD FOR NOISE FILTERING

The method that we propose is based in a fact that if the central element is higher than its neighbors, then these neighbors will get contribution from the central element, whereas in the opposite case, the central will gain the contribution from the neighbors. The proposed methodology flowchart is shown in figure 1.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 14 (2015) pp 33976-33980

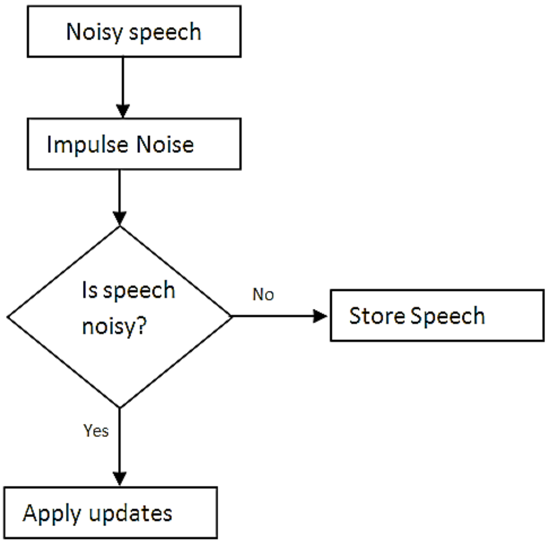


Figure 1: Flowchart of the proposed algorithm

Algorithm:

1. Read the input speech signal I affected by noise.
2. Compute S_{max} , S_{min} , and S_{med} the maximum, minimum and median values of the speech sample in the moore neighborhood where $r = 1$and the central sample of the neighborhood is the sample which is testing for impulse.
3. If $S_{min} < S_{i,j}^t < S_{max}$ then the sample $S_{i,j}^t$ is un-corrupted and its value is unchanged.
4. If the testing speech sample falls in any one of the following category then go to step 5.

a. Test speech sample value is less than all other speech sample values in the neighborhood.

b. Test speech sample value greater than all other sample values in the neighborhood

5. $S_{i,j}(t+1) =$

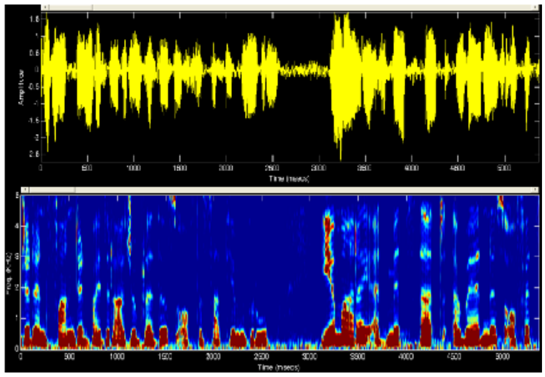
$$\text{Mean}\left(\begin{array}{c} S_{i-1,j-1}(t), S_{i+1,j}(t), \\ S_{i-1,j+1}(t), S_{i,j-1}(t), \\ S_{i,}(t), S_{i,j+1}(t), \\ S_{i+1,j-1}(t), S_{i+1,j}(t), S_{i+1,j+1}(t) \end{array}\right)$$

6. Repeat the step 2 to 4 for all the speech sample of input speech I.

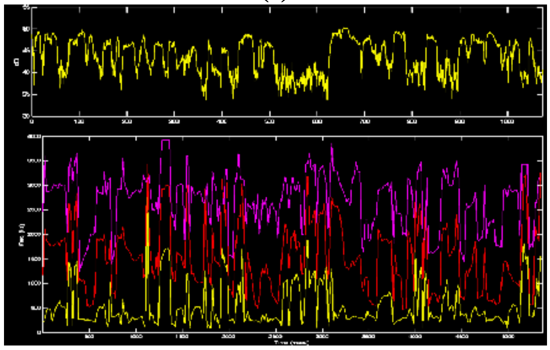
The major advantage of this method is that the speech signal is first checked for impulse noise and only then it is replaced, otherwise it is left unchanged. We can improve the performance by using non-uniform C.A. rules. As the output of the given method looks moderate at low noise ratio up to 40% but shows poor performance as the noise ratio increases. Also, we can find a condition to determine the quantity of iterations needed to filter a speech signal without leave undesirable noises and without endanger the specific speaker features of interest.

Figure 2(a) is a standard speech signal & and spectrogram recorded for training and testing for speaker recognition.

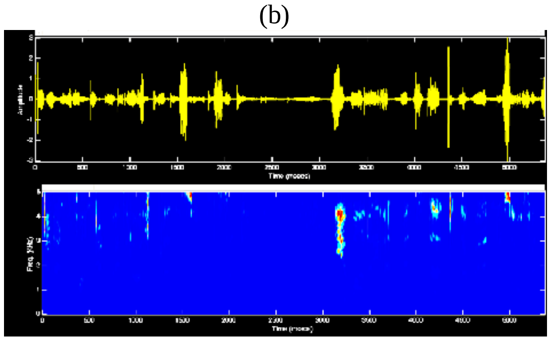
Figure 2(b) is energy and formants track of the standard signal. Figure 2(c) is speech signal after the application of cellular automata based noise filtering and spectrogram. Figure 2 (d) is energy and formants track of the given speech signal after application of cellular automata filtering.



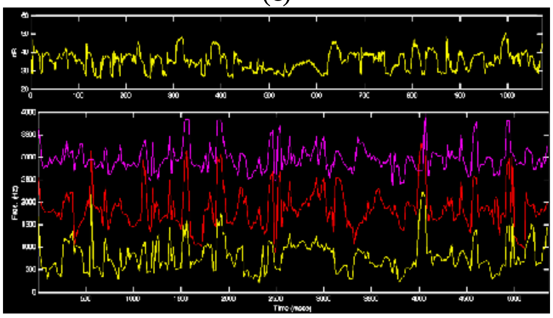
(a)



(b)



(c)



(d)

Figure. 2: (a) Original signal of Speech with noise and spectrogram (b) Energy and formants track of noisy speech signal (c) illustrates the resulted speech signal and spectrogram after filtered using CA. (d) illustrates the resulted energy and formants track after filtered using CA.

V. SPEAKER RECOGNITION USING CELLULAR AUTOMATA BASED FILTERED SPEECH SIGNAL

According to the psychophysical studies, the auditory system of human being responses logarithmically in a nonlinear scale known as Mel scale [11, 12]. The mel scale approximates better the auditory system's of human being as compare to frequency bands which are spaced linearly.Mel Frequency Cepstrum Coefficients (MFCC), which use the Mel scale, are the most commonly used acoustic features for speaker recognition. Traditional MFCC based S-R-S (speaker-recognition-system) uses triangular filter [13].

The triangular filter response $\Psi_i(k)$ of i^{th} filter is

$$\Psi_i(k) = \begin{cases} 0 \text{ for } k \le k_{bi-1} \\ \dfrac{k - k_{bi-1}}{k_{bi} - k_{bi-1}} \text{ for} k_{bi-1} \le k \le k_{bi} \\ \dfrac{k_{bi+1} - k}{k_{bi+1} - k_{bi}} \text{ for} k_{bi} \le k \le k_{bi+1} \\ 0 \text{ for } k \ge k_{bi+1} \end{cases} \quad (5)$$

In this work, Gaussian filters (GF) and Tukey filters have been used as averaging bins inplace of triangular filters for calculating MFCC as in a typical speaker recognition application. Gaussian and Tukey filters can provide much smoother transition from one sub band to other, preserving most of the correlation factors. The mean and variance of a GF and Tukey filter can be independently chosen and this gives the advantage of gaining control over the amount of overlaps with neighboring sub bands. Also, the design parameters given for GF filter and Tukey filter can also be calculated from mid point and also from end-points which are present at the base of the TF (Triangular filters) used for MFCC. The expression for GF is written as [14]

$$\Psi_i^{gMFCC} = e^{-\frac{(k-k_{bi})^2}{2\sigma_i^2}} \quad (6)$$

Based on the evidence Tukey filter is a combination of the rectangular window and the Hann window [15]. In fact it is a cosine-tampered window and is defined as follows

$$\Psi_i(k) = \begin{cases} \dfrac{1}{2}\left(1 - \cos\left(2\pi.\dfrac{k - k_{bi-1}}{N_{Hann}}\right)\right), \\ \text{for } k_{bi-1} \le k_{bi-1} + 1 \le k_{bi-1} + \dfrac{N_{Hann}}{2} \\ 1, \\ k_{bi-1} + \dfrac{N_{Hann}}{2} + 1 \le k \le k_{bi+1} - \dfrac{N_{Hann}}{2} \\ \dfrac{1}{2}\left(1 - \cos\left(2\pi.\dfrac{k - N_{Rect}}{N_{Hann}}\right)\right), \\ k_{bi+1} - \dfrac{N_{Hann}}{2} \le k \le k_{bi+1} \\ 0, \text{for } k \ge k_{b\,i+1} \end{cases} \quad (7)$$

Fuzzy Vector Quantization (FVQ) technique uses the principle that any feature vector present between the two or more clusters should not be assigned to any one cluster. Therefore in FVQ any feature vector will have association with all clusters [16]. Fuzzy c-means, it is a clustering technique that permits one data to belong to more than one cluster at the same time.

In this work, clean representation of a speech sample is obtained using cellular automata based filtering. About 100 samples (TIMIT and self-collected) have been collected and to this applied cellular automata based filtering to speech samples, MFCC and FVQ have been used as in traditional speaker recognition techniques.

The filters discussed above are considered and for each of the case, the accuracy of recognition has been calculated. These results are discussed in the following section.

VI.EXPERIMENTAL RESULTS

The performance of an MFCC based classifier has been evaluated where each feature set was tested using TF, GF as well as Tukey Filter. A total of 800 utterances were put to test. For the above cases, identification / recognition accuracy was calculated using the expression:

$$\text{Percentage of Identification Accuracy} = \frac{\text{No of utterance correctly identified}}{\text{Total No of utterance under test}}$$

Table 3 shows the identification accuracies for the sparse representation of TIMIT database for TF, GF and Tukey based filters respectively. It can be observed from this table that use of GF shows significant improvement upto 96%.

Table 3: Speaker Recognition Accuracy (%) of TIMIT Database

No of Speakers	Triangular Filter	Gaussian Filter	Tukey Filter
100	95	96	95.5

Table 4 shows the identification accuracies for self-collected database for TF, GF and Tukey based filters respectively. It can be observed from this table that GF shows significant improvement up to 94%.

Table 4: Speaker Recognition Accuracy (%) of Self Collected Database

No of Speakers	Triangular Filter	Gaussian Filter	Tukey Filter
100	93	94	93.5

Table 5 shows the identification accuracies for the CA filtered speech of TIMIT Database for TF, GF and Tukey based filters respectively. It can be observed from this table that GF shows 98% of accuracy for database size of 100 speakers. The performance of GF based speaker recognition is high compared to TF and Tukey filter.

Table 5: Speaker Recognition Accuracy (%) using CA filtered speech of TIMIT Database

No of Speakers	Triangular Filter	Gaussian Filter	Tukey Filter
100	96	98	97

Table 6 shows the identification accuracies for the CA filtered speech of self-collected database for TF, GF and Tukey based filters respectively. It is observed from this table that GF shows 96% of accuracy for database size of 100 speakers.

Table 6: Speaker Recognition Accuracy (%) using speech of Self collected Database

No of Speakers	Triangular Filter	Gaussian Filter	Tukey Filter
100	94	96	95

VII.CONCLUSIONS

Efficiency of the speaker recognition system proposed in this research, on TIMIT database is 98% and self collected database is 96%. This system can be used in civil applications as well as in military applications where large data is required to store information of signals that is voice samples.

REFERENCES

[1] J. V. Neumann, "Theory of Self-Reproducing Automata", (University of Illinois Press), 1966.

[2] S.Wolfram, "Computation Theory of Cellular Automata ", Comm.Math. Phys., 1984, 15-57.

[3] R.C.Gonzales and R.E.Woods R, "Digital Image Processing", 2nd ed. Reading, (MA: Addison – Wesley), 2002.

[4] I.Pitas, A.Venetsanopou, "Nonlinear Digital Filters: Principles and Application", Norwell, MA: Kluwer, 1990.

[5] T. S. Huang, G.J.Yang, and G.Y.Tang, "Fast two dimensional median filtering algorithm", IEEE Transactions on Acoust. Speech Signal Process.27(1), 1997, pp.13–18.

[6]. H. Wang and R.A.Hadad, "Adaptive median filters: new algorithms and results", IEEE transactions on image processing, 4(4), 1995, pp.499-502.

[7]. Z. Wang and D.Zhang," Progressive switching median filters for the removal of impulsive noise from highly corrupted images", IEEE transactions on circuits and system II: analog and digital signal processing, 46(1), 1999, pp.78-80.

[8]. A.Popovici and D.Popovici, "Cellular automata in image processing", In Proceedings of the 15th International Symposium on the Mathematical Theory of Networks and Systems, D. S. Gilliam and J. Rosenthal, electronic proceedings. Eds., 2002.

[9]. M.A. Peer, N.A.Shah and R.A. Khan," Image processing systems based on cellular automata", proceeding of the national conference on recent advances and future in IT, RAFIT, 2005.

[10] L.Songtao, C.Hongguan, Y.Shaoging, "An effective filtering algorithm for salt-peper noises based on cellular automata", IEEE congress on image and signal processing,2008.

[11] T.Ganchev, N.Fakotakis and G. Kokkinakis, "Comparative Evaluation of Various MFCC Implementations on the Speaker Verification Task", Proc. of SPECOM, Patras, Greece, 2005, pp. 1191-1194.

[12] J. Daniel, Mashao, S.Marshalleno, "Combining Classifier Decisions for Robust Speaker Identification: Pattern Recognition", vol. 39, 2006, pp. 147-155.

[13] S. Chakroborty, and G.Saha, "Improved Text-Independent Speaker Identification using Fused MFCC &IMFCC Feature Sets based on Gaussian Filter",World Academy of Science, Engineering and Technology, 2009, Vol:3, pp. 11-26.

[14] A.V. Oppenheim, R. W. Schafer, J. R. Buck, "Discrete-Time Signal Processing", 2nd ed., Upper Saddle River,NJ, Prentice Hall, 1999.

[15] H.S. Jayanna, S. R. M. Prasanna, "Fuzzy vector quantization for speaker recognition under limited data conditions", TENCON-IEEE Region 10 Conference,2008, pp. 1 – 4.

[16] H. Wang, X. Zhang, H. Suo, Q. Zhao and Y. Yan, "A novel fuzzy-based automatic speaker clustering algorithm ", ISNN,, 2009, pp. 639–646.