ENHANCED SPEAKER RECOGNITION USING INTEGRATED INVERSE WAVE TRANSFORMATION AND LOW PASS FILTER BASED ALGORITHM

Varinderjit Singh[1], Dr Paramjeet Singh[2]
Dept. of Computer Science& Engineering,
GZS PTU CAMPUS, Punjab. India.
[1]vickyss550@gmail.com, [2]param2009@yahoo.com

Abstract

This research work focuses on improving the performance of speaker recognition method by incorporating low pass filter with inverse wave transformation methods. The principle objective is to identify speakers in efficient manner by using the correlation thresholding. Speaker recognition is attractive standard in real time security organizations to improve the unauthorised person's recognition i.e. it can differentiate unwanted persons straightforwardly. The methodologies established so far are running correctly and performs effectively in real time systems but provide poor results in case of the noisy audios. To attain the objectives of this research work, a new hybrid procedure is projected which has the ability to identify the speaker efficiently even in case of noisy signals. The comparison among the proposed and existing techniques has shown that the proposed algorithm performs quite effectively.

Index Terms: **Speaker recognition, audio signals, Inverse Wave transformation, Low pass filter.**

I. Introduction

The speaker distinguishment is the procedure of taking the spoken word as an info and matches it with the database of awhile ago recorded talks on foundation of different parameters. This could be carried out by different techniques. In this exploration work opposite wave conversion is utilized and speaker is distinguished on support of different parameters. A definitive point of Speaker distinguishment examination is to permit a machine to distinguish matches of sound with 100% exactness that are spoken by any individual, autonomous of vocabulary size, clamor, stress, or channel conditions. Despite a few decades of examination here precision more terrific than 90% is just accomplished when the errand is obliged in some structure. Contingent upon how the undertaking is compelled, distinctive levels of execution could be achieved; for instance sound of an individual might be distinguished if it is the same individual or not. For speaker distinguishment of diverse speakers on premise of certain characteristics, precision is not more amazing than 87%, and preparing can take many times constant.

Speaker distinguishment [1] is a methodology of immediately distinguishing who is talking on the foundation of characteristics of speaker of the discourse indicator. Fundamentally, speaker distinguishment is ordered into speaker distinguishment and speaker check. Wide provision of speaker distinguishment framework incorporates control access to administrations, for example, keeping money by phone, database access administrations, voice dialling phone shopping so on. Right away, speaker distinguishment engineering is the most suitable innovation to make new

administrations that will make our regular lives more secured. [1] Speech expressions from one speaker may fluctuate because of age, sex, clamor and nature's domain, and speaker tone and state of mind. Fifty years later of exploration of discourse transforming in the time dominion, the correctness of discourse recognizers has not arrived at an attractive triumph rate. The many-sided quality of speaker distinguishment frameworks has expanded these days.

The normal issue with distinguishment framework these days is that the framework can effortlessly be tricked. Despite the fact that it utilizes biometric distinguishment which is one of a kind from other people, there are still approaches to trick the framework. As for unique mark distinguishment, it doesn't have a great mental impact on the individuals on account of its wide use in wrongdoing examinations. Likewise, when the surface of human unique finger impression is harmed, the distinguishment framework will have issues to distinguish the client in light of the fact that the framework distinguishes the surface of the fingerprints while for face distinguishment, individuals are even now working the posture and the enlightenment invariance

Discourse is a common [2] mode of correspondence for individuals. All the applicable abilities are scholarly

throughout unanticipated adolescence, without direction, and we keep on relying upon discourse correspondence all around our lives. It comes so commonly to us that we don't understand how perplexing a marvel discourse are in [2]. The human vocal tract and articulators are living organs with nonlinear lands, whose operation are under cognizant control as well as influenced by components going from sex to childhood to passionate state. Thus, vocalizations can shift generally as far as their intonation, articulation, enunciation, unpleasantness, nasality, pitch, volume, and rate; also, throughout transmission, our spasmodic discourse examples could be further misshaped by foundation clamor and echoes, and in addition electrical aspects (if phones or other electronic gear are utilized). All these wellsprings of variability make speaker distinguishment exceptionally intricate issue significantly more than discourse era. Yet individuals are so agreeable with discourse that we might likewise want to interface with our machines through discourse, as opposed to utilizing primitive interfaces, for example, consoles and indicating gadgets.

Speaker distinguishment [3] is one of the biometric systems for security. Biometrics is seen by numerous scientists as an answer for a ton of client distinguishment and security issues. This may incorporate speaker distinguishment, face distinguishment, finger impression distinguishment, finger geometry, hand geometry, iris distinguishment, vein distinguishment, and voice and mark distinguishment. Various systems are accessible in literary works to determination the programmed speaker distinguishment issue. For the most part, discourse indicator is stirred up with commotion sign. In any case, a large portion of the work on discourse preparing for the speaker distinguishment was carried out by centering the discourse under the quiet situations and just few by centering discourse under uproarious conditions.

Speaker distinguishment characteristics has a percentage of the focal points like discourse info is not difficult to perform on the grounds that it doesn't oblige a specific expertise as does writing or pushbutton operations. Data could be enter actually when the client is moving or doing different exercises including the hands, legs, eyes, or ears. Since a receiver or phone might be utilized as a data terminal, inputting data is sparing with remote inputting equipped for being achieved over existing phone systems and the Internet.

II. CLASSIFICATION OF SPEAKER RECOGNITION SYSTEMS

Most speaker recognition systems can be classified according to the following categories [4]:

A. Speaker Dependent vs Speaker Independent

A speaker-dependent speaker distinguishment framework is one that is prepared to distinguish the discourse of stand out speaker. Such frameworks are custom manufactured for only a solitary individual, and are consequently not industrially practical. Then again, a speaker-independent framework is one that is freedom is tricky to attain, as speaker distinguishment frameworks have a tendency to get sensitive to the speakers they are prepared on, bringing about slip rates that are higher than speaker subordinate framework.

B. Isolated vs. Continuous

In disengaged discourse, the speaker stops without further ado between every expression, while in consistent discourse the speaker talks in a persistent and conceivably long stream, with practically no breaks in the middle of. Secluded speaker distinguishment frameworks are not difficult to raise, as it is irrelevant to figure out where one expression closures and an alternate begins, and each one saying has a tendency to be all the more neatly and plainly spoken. Words spoken in consistent discourse then again are subjected to the co-verbalization impact, in which the elocution of an expression is altered by the words encompassing it. This makes preparing a discourse framework challenging, as there may be numerous conflicting articulations for the same word.

C. Keyword based vs. Subword unit based

A speaker distinguishment framework could be prepared to distinguish entire words, for instance pooch or feline. This is helpful in provisions like voice-command-systems, in which the framework require just distinguish a little set of words. This methodology, while straightforward, is sadly not versatile. As the concordance of distinguished words develop, so excessively the intricacy and execution time of the recognition.

III. Literature Review

Advancement of speaker distinguishment frameworks started as unanticipated as the 1960s with investigation into voiceprint examination, where aspects of a singular's voice were thought to have the ability to portray the uniqueness of a singular much like a finger impression. The unanticipated frameworks had numerous imperfections and examination resulted to determine a more solid strategy for anticipating the correspondence between two sets of discourse articulations. Speaker distinguishment examination proceeds today under the domain of the field of computerized indicator transforming where numerous developments have occurred lately.

Salomon .J et al (2008) [1] examined the utilization of recurrence changes and example distinguishment to enhance the correctness of single speaker different word speaker distinguishment frameworks. It was nearly identified with indicator subspace based discourse upgrade plans. Rather than universal front-end mixture ward characteristic conversions, where characteristic arrangement is performed utilizing the most noteworthy scoring mixture, the proposed conversion is incorporated inside the speaker distinguishment framework utilizing a probabilistic characteristic arrangement strategy, which invalidates the requirement for recovering the features/retraining the Universal Background Model (UBM).

Addou .D.et.al(2011) [2] depicted a commotion powerful Distributed Speaker distinguishment (DSR) front-closure utilizing a blending of customary Mel-cepstral Coefficient (MFCC) and Line Spectral Frequencies (LSF). These characteristics are satisfactorily changed and diminished in a multi-stream plan utilizing Karhunen-Loeve Transform (KLT).

Gonzalez.j.et.al (2011) [3] introduced a characteristic payment schema dependent upon least mean square slip (MMSE) estimation and stereo preparing information for strong speaker distinguishment with a specific end goal to get clean characteristic assessments. The discrete nature of characteristic space characterization presented significative focal points. It permitted the usage of an exceptionally productive MMSE estimator as far as correctness and computational expense.

Lu.y.et.al (2003) [4] displayed another model adjustment calculation utilizing piecewise straight change (PLT) for hearty speaker distinguishment. In proposed calculation, the nonlinear relationship between preparing and testing mean vectors was approximated by a set of piecewise straight conversions. The PLT coefficients were evaluated from acclimatization information by the desire amplification (EM) calculation and most extreme probability (ML) paradigm. The proposed calculation could defeat the impediment of straight supposition in customary convert based acclimatization calculation.

Alkhaldi.w.et.al (2002) [5] has displayed Discrete Wavelet Transform- based characteristic extraction strategy for multi-band programmed speech/speaker distinguishment. This method has tantamount execution with customary

strategy. In the event that has been found that both strategies are corresponding under jumbled conditions, if the characteristics concentrated utilizing each of them are consolidated.

Wassner. H.et.al (2009) [6] enhanced distinguishment rate by optimisation of Mel Frequency Cepstral Coefficients alterations which concerned the time-recurrence representations to gauge coefficients. There are numerous approaches to get a range out of sign which vary in the system itself (Fourier, Wavelets and so on.), and in the normalisation. It had been demonstrated that clamor safe cepstral coefficients were gotten, for speaker autonomous associated word distinguishment.

Chia. S.et.al (2013) [7] proposed a corruption model which speaks to the ghostly changes of discourse sign expressed in uproarious situations. The model utilized recurrence twisting and abundancy scaling of every recurrence band to reproduce the varieties of formant area, formant data transmission, pitch, otherworldly tilt, and vigor in every recurrence band by Lombard impact. An alternate Lombard impact, the variety of in general vocal power is spoken to by a multiplicative steady term hinging upon ghastly size of info discourse.

Rusyn.b.et.al (2008) [8] showed that Wavelets are capacities that fulfill certain numerical necessities and are utilized within speaking to information or different capacities. In wavelet examination the scale that we use to take a gander at information assumes an exceptional part. Wavelet calculations process information at distinctive scales or resolutions, underlining information's terrible or little characteristics. Nature of division such a great amount of relies on upon phoneme creation of discourse sign.

Daqrouq .K.et.al. (2009) [9] displayed a thought of clamor undoing for the discourse indicate so vigor of the speaker distinguishment framework expanded. Two pieces are there: Discrete Wavelet Transform DWT and Adaptive Linear Neuron (Adaline) Enhancement Method (DWADE) and Wavelet Gender Discrimination (WGD) and Speaker Recognition utilizing Discrete Wavelet Transform (DWT) Power Spectrum Density (PSD).

Gaikwad.s.et.al (2010) [10] gave a diagram of real mechanical view and valuation for the crucial advancement of speaker distinguishment and likewise gives review strategy created in each one phase of speaker distinguishment. This paper helps in picking the procedure plus their relative benefits & faults.

Waelal.s.et.al (2009) [11] proposed a framework that discovers the right character of speaker on support of

Continuous, Discrete Wavelet Transform and Power Spectrum Density. The framework hinges on upon the multi-stage characteristics removing because of its better correctness. Great proficience was demonstrated by frameworks dependent upon multistage characteristic following. The impact of Wavelet Transform on speaker characteristic concentrating was contemplated.

IV. DATABASE

The database comprises of recorded voices which are utilized for matching with the information tests. This database comprises of 100 separate voices. These are diverse pictures with distinctive postures from distinctive kind of sources.

Different specimens are taken in which 100 set of distinguished persons and 100 set of unrecognized persons is brought and after that tried with the beforehand archived 10 guaranteed distinguished voices.

The Table 1 shows the amount of guaranteed sound examples that are spared consistent with which the inputs could be matched. For the trial setup, 10 examples of confirmed sound cuts are taken. 100 specimens of unrecognized inputs are taken which must be matched with affirmed sound examples. Again 100 specimens of obscure individual sound recordings are taken.

Table 1: Set of Samples

Type	Sample Size
Certified	10
Input Recognized	100
Unknown Person Audio	100

The 200 examples in aggregate are taken for testing. For the specimens are taken then hits are computed which gives the amount of times the right distinguishment is carried out when the info is available in the awhile ago spared set of guaranteed voices. Distinguishment is carried out by matching the info examples voices having wave design with the beforehand guaranteed set of specimens. Likewise, number of miss are assessed appropriately voices when example is not found in the record of guaranteed voices.

The proposed calculation has utilized method of reverse wave conversion and the different steps included in the calculation are talked about in this part. The transpose is taken with the goal that the discourse outlines that are for the most part indicated on a level plane could be changed over in vertical structure so it might be straightforward and distinguish the discourse outlines effortlessly. Matlab 7.10 has been utilized for the execution of converse wave change calculation. 100 specimens of regarded as well as obscure individual's voice examples have been brought and matched with 10 confirmed voice tests.

V. Performance Analysis

The execution of the proposed method is superior to the existing methods. The greater part of existing strategies does distinguishment of discourse through the content wrapping. These methods take an excess of time in distinguishing the discourse. These procedures are not precise in the nature. As the phonemes must be distinguished utilizing neural systems. Neural systems take an excessive amount of time in preparing and testing. Neural systems can promptly be connected to static or transiently limited example distinguishment undertakings however can't be effectively connected to element and transiently broadened example distinguishment assignments. In this manner, in a speaker distinguishment framework, it at present bodes well for use neural systems for acoustic modelling, not for worldly modelling. In the proposed calculation discrete cosine converts based sound packing and channels are utilized for correctness and the matching is carried out utilizing opposite wave changes which diminish the time for distinguishment of voices.

The 100 specimens in sum are taken for testing. For the examples are taken then hits are figured which gives the amount of times the right distinguishment is carried out when the data is available in the formerly spared set of guaranteed voices. Distinguishment is carried out by matching the data examples voices having wave position with the long ago guaranteed set of specimens. Correspondingly, number of miss are assessed appropriately voices when example is not found in the record of ensured voices.

At that point as per equations correctness and slip rates are figured .Firstly precision and mistake rates are computed in uproarious environment when channel and DCT are not utilized within sound distinguishment.

The point when the amount of information examples is 10 then there are 2 miss and 8 hits so the correctness rate is 80 percent and slip rate is 20 percent. The point when the amount of data specimens is expanded to 20 then there is 4 miss and exactness rate is 80 percent so the blunder rate is 20 percent. The point when again the amount of info specimens expansions to 30 then there are 6 miss and 24 hits so the correctness rate is 80 percent and blunder rate is 20. After that again when specimen number expanded up to 90 then the amount of miss expansions to 16 and lapse rate is 18 so the correctness rate turns into 82. The point when tests

taken are 100 then there is just 17 miss so the correctness rate is 83 and mistake rate turns into 17.

Table 2 has shown the comparative analysis of Hits between the proposed and existing speaker recognition algorithms. It has clearly shown that the no of hits are more in case of the proposed algorithms therefor it has shown quite significant improvement over the existing algorithm.

Table No. 2 Comparative analysis of Hits

No. of audio signals	Existing algorithm	Proposed algorithm
10	8	9
20	17	18
30	25	26
40	33	35
50	41	45
60	50	54
70	58	64
80	66	73
90	76	82
100	84	91

Table No. 3 Comparative analysis of misses

No. of audio signals	Existing algorithm	Proposed algorithm
10	2	1
20	3	2
30	5	4
40	7	5
50	9	1
60	10	6
70	12	6
80	14	7
90	14	8
100	16	9

Table 3 has shown the comparative analysis of Misses between the proposed and existing speaker recognition techniques. It has clearly publicised that the no of misses are minimum in case of the proposed algorithms than the existing therefor it has shown quite significant improvement over the existing algorithm.

Table No. 4 Comparative analysis of accuracy (%)

No. of audio signals	Existing algorithm	Proposed algorithm
10	80	90
20	85	90
30	83	88
40	82	87
50	81	90
60	83	91
70	82	90
80	82	90
90	83	91
100	84	91

Table 4 has shown the comparative analysis of accuracy between the proposed and existing speaker recognition algorithms. It has clearly shown that the accuracy is more in case of the proposed algorithms therefor it has shown quite significant improvement over the existing algorithm.

Table No. 5 Comparative analysis of error rate (%)

No. of audio signals	Existing algorithm	Proposed algorithm
10	20	10
20	15	10
30	17	12
40	18	13
50	19	10
60	17	9
70	18	10
80	18	10
90	17	9
100	16	9

Table 5 has shown the comparative analysis of error rate between the proposed and existing speaker recognition techniques. It has clearly publicised that the error rate is minimum in case of the proposed algorithms than the existing therefor it has shown quite significant improvement over the existing algorithm.

The point when the amount of info specimens is 10 then there is just 1 miss and 9 hits so the correctness rate is 90 percent and slip rate is 10 percent. The point when the amount of information examples is expanded to 20 then there is again 1 miss so the precision rate is 95 percent and lapse rate is just 5 percent. The point when again the amount of info examples increments to 30 then there are 2 miss and 28 hits so the exactness rate is 93 percent and lapse rate ascents to 7. After that again when specimen number expanded up to 90 then the amount of miss increments to 9 and blunder rate likewise expands to 10 so the precision rate turns into 90. The point when tests taken are 100 then there is just 9 miss so the precision rate is 91 and slip rate turns into 9.

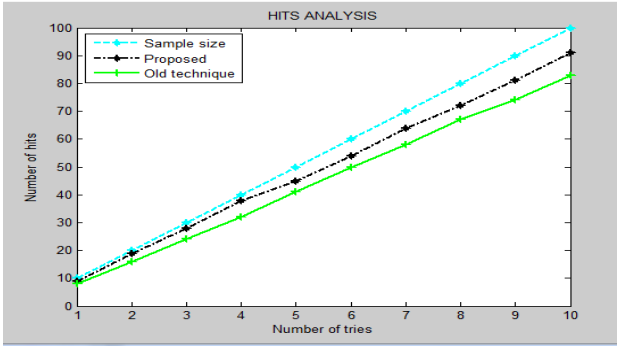

Figure 1: Hits Analysis

The Figure 1 shows to the hits investigation. It demonstrates the distinction between the amount of hits of existing methods and proposed system. So as to recognize them distinctive colours has been utilized. In Figure 1 X-pivot speaks to the amount of tries and Y-hub speaks to the

amount of hits. It has been plotted with diverse qualities of data examples and hits are computed from the distinguishment of addresses. It is obviously seen that the hit proportion of old system is quite low as contrasted with new strategy. Subsequently, more hits are acquired utilizing DCT as a part of speaker distinguishment.

The Figure 2 shows to the miss investigation. It indicates the contrast between the amount of misses of existing strategies and proposed procedure. In Figure 2 X-hub speaks to the amount of tries and Y-hub speaks to the amount of misses. It has been plotted with distinctive qualities of data specimens and misses are figured from the distinguishment of discourses. It is unmistakably seen that the miss degree of new system is quite low as contrasted with old procedure. Therefore, less misses are acquired utilizing DCT within speaker distinguishment.

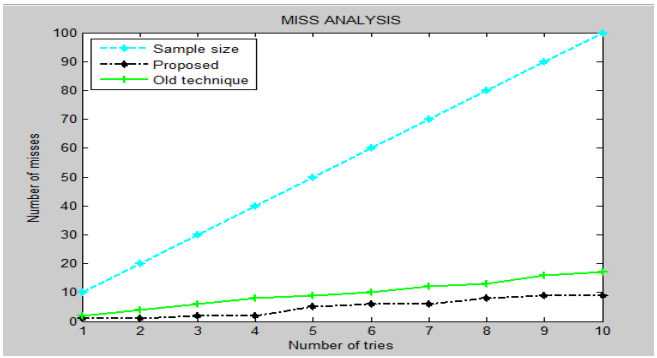

Figure 2: Miss Analysis

The Figure 3 speaks to the correctness examination. It shows the contrast between the correctness of existing methods and proposed method. In Figure 3 X-hub speaks to the amount of tries and Y-hub speaks to the exactness in rate. The most extreme worth of correctness is 100 percent. It has been plotted with distinctive qualities of data specimens and exactnesses are figured from the distinguishment of discourses. It is obviously seen that the precision of new method is quite high as contrasted with old strategy. Accordingly, higher correctness is gotten just when there are more hits and lesser misses. It is plainly seen that precision of the old systems is quite low as near X-hub and the correctness of the new system is quite high as near greatest quality. In proposed method channel and DCT are utilized with opposite wave change. Discrete cosine transform based sound squeezing is utilized to lessen the span of sound indicators and evacuate clamour from the sound signs.

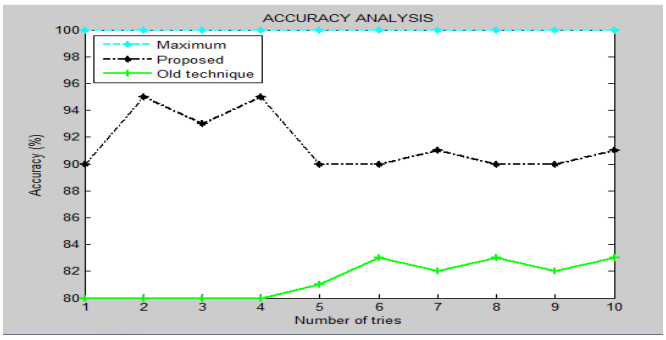

Figure 3: Accuracy analysis

The Figure 4 signifies the error rate investigation. It shows the difference among the error rate of existing methods and projected method. In Figure 4 X-axis signifies the number of tries and Y-axis characterize error rate in percentage. The supreme value of error rate is 100 percent. It has been plotted with different values of input samples and error rates are calculated from the recognition of speeches. It is clearly seen that the error rate of new technique is very low as compared to old technique. As a result, less error rate is obtained only when there are more hits and lesser misses. It is clearly seen that error rate of the old techniques is very high as close to the maximum value of error rate and the error rate of the new technique is very low as close to X-axis. Error rate is lower when the hits are more and error rate rises with rise in miss ratio.

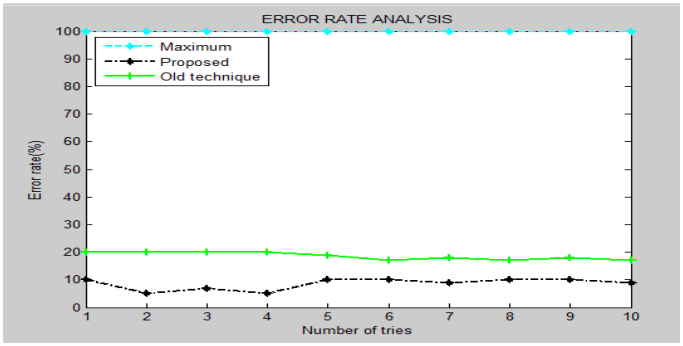

Figure 4: Error Rate Analysis

Conclusion and Future Work

This research work has projected new speaker identification algorithm which is significant in refining the speaker recognition performance. The proposed technique is intelligent to validate the specific speaker based on the individual info that is comprised in the signal and the appreciation is completed using converse wave alteration. As the voice of specific person is dissimilar from another in terms of pitch and other potentials so one can easily discriminate the speeches of dissimilar persons. So the proposed technique of reverse wave transformation assesses the variation in two dissimilar speeches by interpretation of the speech waves and calculates the correlation. The correlation is then evaluated using correlation thresholding to identify the speaker. The results have shown that planned technique provides high accuracy rate.

In near future we will extend this work by using the neural networks to evaluate the correlation. However some other filters will also be used to improve the accuracy rate further.

References

[1] Jorge Salomon Fuentes, Dr. Chit-Sang Tsang "Speaker recognition using Frequency Transformations", 978-1-4244-2622-5/09, IEEE, 2009.

[2] D. Addou, S.A. Selouani, M. Boudraa, B. Boudraa "Transform-based multi-feature optimization for robust distributed speaker recognition" IEEE GCC conference and exhibition, februaury, 2011.

[3] José A. González, Antonio M. Peinado, Angel M. Gómez, and José L. Carmona "Efficient MMSE Estimation and Uncertainty Processing for Multi-environment Robust Speaker recognition" IEEE transactions on audio, speech, and language processing, vol. 19, no. 5, JULY 2011,

[4] Yong Lü and Zhenyang Wu "Maximum Likelihood Model Adaptation Using Piecewise Linear Transformation for Robust Speaker recognition" The 13th IEEE International Symposium on Consumer Electronics ISCE2009.

[5] W. Alkhaldi, W. Fakhr and N. Hamdy, "Automatic Speech/Speaker Recognition In Noisy Environments Using Wavelet Transform" IEEE 2002.

[6] Hubert Wassner I, Gerard Chollet "New Cepstral Representation Using Wavelet Analysis And Spectral Transformation For Robust Speaker recognition" 2009.

[7] Sang-Mun Chi, Yung-Hwan Oh "Lombard Effect Compensation And Noise Suppression For Noisy Lombard Speaker recognition" 2013.

[8] Bohdan Rusyn, Andriy Chorniy "Application of Wavelet-Transformation in to the System of Speaker recognition" TCSET'2008, February 19-23, 2008.

[9] K. Daqrouq, T. Abu Hilal, M. Sherif, S. El-Hajjar, and A. Al-Qawasmi "Speaker Verification System Using Discrete Wavelet Transform And Formants Extraction Based On The Correlation Coefficient " ; proceeding of international multi conference of engineers and computer scientists, IMECS 2011 .

[10] Santosh K.Gaikwad, Bharti W.Gawali, Pravin Yannawar "A Review on Speaker recognition Technique" International Journal of Computer Applications (0975 – 8887) Volume 10– No.3, November 2010.

[11] Wael al-sawalmeh, Khaleddaqrouq, Abdel-Rahman Al-Qawasmi, and Tareq Abu Hilal "The use of wavelets in speaker feature tracking recognition System Using Neural Network" , Vol 5, ISSN: 1790-5052 WSEAS TRANSACTIONS ON SIGNAL PROCESSING, May 2009.

